Tarek Sobh · Khaled Elleithy
Ausif Mahmood
Mohammad A. Karim
Editors

Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics

Springer

Novel Algorithms and Techniques in Telecommunications,
Automation and Industrial Electronics

Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics

Edited by

Tarek Sobh
Khaled Elleithy
Ausif Mahmood
Mohammad A. Karim

Springer

Editors
Dr. Tarek Sobh
University of Bridgeport
School of Engineering
221 University Avenue
Bridgeport CT 06604
USA
sobh@bridgeport.edu

Prof. Khaled Elleithy
University of Bridgeport
School of Engineering
221 University Avenue
Bridgeport CT 06604
USA
elleithy@bridgeport.edu

Ausif Mahmood
University of Bridgeport
School of Engineering
221 University Avenue
Bridgeport CT 06604
USA

Prof. Mohammad A. Karim
Old Dominion University
2035 Hughes Hall
Norfolk, VA 23529
USA

ISBN: 978-1-4020-8736-3 e-ISBN: 978-1-4020-8737-0

Library of Congress Control Number: 2008932235

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Contents

Preface

This book includes the proceedings of the 2007 International Conference on Telecommunications and Networking (TeNe) and the 2007 International Conference on Industrial Electronics, Technology &Automation (IETA).

TeNe 07 and IETA 07 are part of the International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering (CISSE 07). The proceedings are a set of rigorously reviewed world-class manuscripts presenting the state of international practice in Innovative Algorithms and Techniques in Automation, Industrial Electronics and Telecommunications.

TeNe 07 and IETA 07 are high-caliber research conferences that were conducted online. CISSE 07 received 750 paper submissions and the final program included 406 accepted papers from more than 80 countries, representing the six continents. Each paper received at least two reviews, and authors were required to address review comments prior to presentation and publication.

Conducting TeNe 07 and IETA 07 online presented a number of unique advantages, as follows:

- All communications between the authors, reviewers, and conference organizing committee were done on line, which permitted a short six week period from the paper submission deadline to the beginning of the conference.
- PowerPoint presentations, final paper manuscripts were available to registrants for three weeks prior to the start of the conference.
- The conference platform allowed live presentations by several presenters from different locations, with the audio and PowerPoint transmitted to attendees throughout the internet, even on dial up connections. Attendees were able to ask both audio and written questions in a chat room format, and presenters could mark up their slides as they deem fit.
- The live audio presentations were also recorded and distributed to participants along with the power points presentations and paper manuscripts within the conference DVD.

The conference organizers and we are confident that you will find the papers included in this volume interesting and useful. We believe that technology will continue to infuse education thus enriching the educational experience of both students and teachers.

Tarek M. Sobh, Ph.D., PE
Khaled Elleithy, Ph.D.,
Ausif Mahmood, Ph.D.
Mohammad A. Karim, Ph.D.
Bridgeport, Connecticut
June 2008

Acknowledgements

The 2007 International Conferences on Telecommunications and Networking (TeNe) and Industrial Electronics, Technology & Automation (IETA) and the resulting proceedings could not have been organized without the assistance of a large number of individuals. TeNe and IETA are part of the International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering (CISSE). CISSE was founded by Professors Tarek Sobh and Khaled Elleithy in 2005, and they set up mechanisms that put it into action. Andrew Rosca wrote the software that allowed conference management, and interaction between the authors and reviewers online. Mr. Tudor Rosca managed the online conference presentation system and was instrumental in ensuring that the event met the highest professional standards. We also want to acknowledge the roles played by Sarosh Patel and Ms. Susan Kristie, our technical and administrative support team.

The technical co-sponsorship provided by the Institute of Electrical and Electronics Engineers (IEEE) and the University of Bridgeport is gratefully appreciated. We would like to express our thanks to Prof. Toshio Fukuda, Chair of the International Advisory Committee and the members of the TeNe and IETA Technical Program Committees including: Abdelshakour Abuzneid, Nirwan Ansari, Hesham El-Sayed, Hakan Ferhatosmanoglu, Ahmed Hambaba, Abdelsalam Helal, Gonhsin Liu, Torleiv Maseng, Anatoly Sachenko, Paul P. Wang, Habib Youssef, Amr El Abbadi, Giua Alessandro, Essam Badreddin, John Billingsley, Angela Di Febbraro, Aydan Erkmen, Navarun Gupta, Junling (Joyce) Hu, Mohamed Kamel, Heba A. Hassan, Heikki N. Koivo, Lawrence Hmurcik, Luu Pham, Saeid Nahavandi, ElSayed Orady, Angel Pobil, Anatoly Sachenko, Sadiq M. Sait, Nariman Sepehri, Bruno Siciliano and Keya Sadeghipour.

The excellent contributions of the authors made this world-class document possible. Each paper received two to four reviews. The reviewers worked tirelessly under a tight schedule and their important work is gratefully appreciated. In particular, we want to acknowledge the contributions of the following individuals: A.B.M. Mozazammel Hossain, Aneesh Chand, Cao Yali, Dioubate Mamady, Eman Abdelfattah, Gayan Hettiarachchi, Grofu Florin, Hatim M tahir, Jing Zhang, K.v.d Kiran Krishnamurthy Ningappa, Kshitij Gupta, Laura Muñoz, Luis Madera, Martin Braae, Muhammad Irfan, Peter Nabende, Pramod Kumar Sharma, Praveen Kumar Kollu, Qiuyun Fu, Radu-Daniel Tomoiaga, Sarhan Musa, Sarosh H. Patel, Shafqat Hameed, Show-Shiow Tzeng, Taner Arsan, Thomas Mundt, Wang Haoping, Yenumula Reddy, and Zornitza Prodanoff

Tarek Sobh, Ph.D., P.E.
Khaled Elleithy, Ph.D.
Ausif Mahmood, Ph.D.
Mohammad A. Karim, Ph.D.

Bridgeport, Connecticut
June 2008

Kernel Locally Linear Embedding Algorithm for Quality Control

Thrasivoulos Tsagaroulis and A. Ben Hamza
Concordia Institute for Information Systems Engineering
Concordia University, Montreal, QC, Canada
{t_tsagar, hamza}@encs.concordia.ca

***Abstract*- In this paper, we introduce a new multivariate statistical process control chart for outliers detection using kernel local linear embedding algorithm. The proposed control chart is effective in the detection of outliers, and its control limits are derived from the eigen-analysis of the kernel matrix in the Hilbert feature space. Our experimental results show the much improved performance of the proposed control chart in comparison with existing multivariate monitoring and controlling charts.**

I. INTRODUCTION

Traditional process monitoring consists of measuring and controlling several process variables at the same time [1]. It is increasing difficult to determine the root cause of defects if multiple process variables exhibit outliers or process deviations at the same moment in time. Multivariate quality control methods overcome this disadvantage by monitoring the interactions of several process variables simultaneously and determining hidden factors using dimensionality reduction [2]. The use of multivariate statistical process control is also facilitated by the proliferation of sensor data that is typically complex, high-dimensional and generally correlated. Complex processes can be monitored the stability evaluated, using multivariate statistical process control techniques.

There are typically two phases in establishing multivariate control charts. The data collected in phase I are used to establish the control limits for phase II.

In recent years, a variety of statistical quality control methods have been proposed to monitor multivariate data including Hotelling's T^2-statistic chart [1], and the principal component analysis control chart based on principal component analysis [4]. These control charts are widely used in industry, particularly in assembly operations and chemical process control [2]. The T^2-statistic is, however, vulnerable to outliers and in order to obtain significantly good results, both the mean and the covariance matrix must be robustly estimated [5–8]. Also, principal component analysis is very sensitive to outliers [2].

In this paper, we present a new multivariate statistical process control chart using kernel locally linear embedding. Locally linear embedding (LLE) is a recently proposed unsupervised procedure for mapping high-dimensional data nonlinearly to a lower-dimensional space [12]. The basic idea of LLE is that of global minimization of the reconstruction error of the set of all local neighborhoods in the data set. The proposed kernel LLE control chart is robust to outlier detection, and its control limits are derived from the eigen-analysis of the kernel LLE matrix in the Hilbert feature space.

The remainder of the paper is organized as follows. Section II briefly reviews some existing multivariate quality control charts. In Section III, we propose a kernel LLE control chart. In Section IV, we demonstrate through experimental results that the performance of the proposed multivariate control chart has greatly been improved in comparison with existing monitoring and controlling charts. Finally, we conclude in Section V.

II. RELATED WORK

In this section, we briefly review some multivariate control charts that are closely related to our proposed approach.

A. Hotelling's T-squared statistic

Let $X = [x_1, x_2, \ldots, x_n]^T$ be an n x p data matrix of n vectors $x_i \in \mathbb{R}^p$, where each observation $x_i = (x_{i1}, \ldots, x_{ip})$ is a row vector with p variables.

Phase I of the T^2 control chart consists of establishing an outlier free reference sample [2]. Hotelling's T^2 statistic, also referred to as Mahalanobis distance, is defined by (1).

$$T_i^2 = (x_i - \bar{x}) S^{-1} (x_i - \bar{x})^T \tag{1}$$

Equations (2) and (3) are the sample mean and covariance matrix respectively.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ and,} \tag{2}$$

$$S = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^T (x_i - \bar{x}) \tag{3}$$

The Phase I upper control limit (UCL) and lower control limit (LCL) of the T^2 control chart are given by (4) and (5).

$$UCL = \frac{(n-1)^2}{n} \frac{[\frac{p}{n-p-1}] F_{\frac{a}{2},p,n-p-1}}{1+[\frac{p}{n-p-1}] F_{\frac{a}{2},p,n-p-1}} \tag{4}$$

$$LCL = \frac{(n-1)^2}{n} \frac{[\frac{p}{n-p-1}] F_{1-\frac{a}{2},p,n-p-1}}{1+[\frac{p}{n-p-1}] F_{1-\frac{a}{2},p,n-p-1}} \tag{5}$$

where F_{a,v_1,v_2} is the value of the inverse of the F cumulative distribution with v_1 and v_2 degrees of freedom, evaluated at the confidence level (1–a).

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 1–6.

In phase II, any outliers identified during phase I are removed and the remaining observations are used to recalculate the T^2 statistic. In other words, the Phase II T^2 statistic is given by (6),

$$T_i^2 = (\tilde{x}_i - \bar{x})S^{-1}(\tilde{x}_i - \bar{x})^T \quad (6)$$

where $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n]^T$ is the new observed data matrix, also referred to as the historical data. Again, any historical data that is plotted outside the control limits and had an assignable cause determined are discarded. Phase II verifies if the process is generating and maintaining values that are considered in control. The control limits for phase II are defined in (7) and (8).

$$UCL = \frac{p(n+1)(n-1)}{n(n-p)} F_{\frac{\alpha}{2},p,n-p} \quad (7)$$

$$LCL = \frac{p(n+1)(n-1)}{n(n-p)} F_{1-\frac{\alpha}{2},p,n-p} \quad (8)$$

Unlike the univariate control charts, the T^2 statistic does not represent the original variables and therefore when an out of control situation occurs we can not determine if it was due to an excess variation of a particular variable or due to a change in the covariance/correlation matrix.

To circumvent these problems, the principal component chart may be used. This control chart can detect changes in the covariance/correlation structure and it may indicate the specific domain that created this excess variation [2]. It also has the advantage of reducing the number of dimensions that need to be analyzed [4].

B. Principal Component Analysis

Principal component analysis (PCA) is a method for transforming the observations in a dataset into new observations which are uncorrelated with each other and account for decreasing proportions of the total variance of the original variables. Each new observation is a linear combination of the original observations.

Standardizing the data is often preferable when the variables are in different units or when the variance of the different columns of the data is substantial. The standardized data matrix is given by (9), where $\mathbf{1} = (1,\ldots,1)^T$ is a n x 1 vector of all 1's, and $D = \text{diag}(S)$ is the diagonal of the covariance matrix.

$$Z = (X - 1\bar{x})D^{-1/2} \quad (9)$$

It is worth pointing out the covariance matrix R of the standardized data Z is exactly the correlation matrix of the original data, and it is given by (10).

$$R = D^{-1/2}SD^{-1/2} \quad (10)$$

PCA is then performed by applying eigen-decomposition to the matrix R, that is $R=A\Lambda A^T$ where $A=(a_1,\ldots,a_p)$ is a p x p matrix of eigenvectors (also called principal components) and $\Lambda=\text{diag}(\Lambda_1,\ldots,\Lambda_p)$ is a diagonal matrix of eigenvalues. These eigenvalues are equal to the variance explained by each of the principal components, in decreasing order of importance. The principal component score matrix is an n x p data matrix Y given by (10), which is the data mapped into the new coordinate system defined by the principal components.

$$Y = ZA = (y_1, ..., y_n)^T \quad (10)$$

Moreover, the covariance of Y is defined in (11).

$$\text{cov}(Y) = \frac{1}{n-1} Y^T Y = \frac{1}{n-1} A^T Z^T ZA = \Lambda. \quad (11)$$

Hence, besides retaining the maximum amount of variance in the projected data, PCA also has the following property: the projected data y_k are uncorrelated with variance equal to $\text{var}(y_k)=\lambda_k$, for $k=1,\ldots,p$.

Assuming we want 99.7% confidence intervals, the upper control limit (UCL), the center line (CL) and the lower control limit (LCL) are given by (12).

$$\begin{aligned} UCL &= +3\sqrt{\lambda_k} \\ CL &= 0 \\ LCL &= -3\sqrt{\lambda_k} \end{aligned} \quad (12)$$

The main drawback of principal component analysis is its sensitivity to outliers [9, 2]. In the next section, we propose a robust multivariate control chart to overcome the problems mentioned above.

III. PROPOSED METHOD

LLE algorithm aims at finding an embedding that preserves the local geometry in the neighborhood of each data point. First, we build a sparse matrix of local predictive weights $W_{i,j}$, such that $\Sigma_j W_{i,j} = 1$, $W_{i,j} = 0$ if x_j is not a k-nearest neighbor of x_i and then $\Sigma_j(W_{i,j}x_j - x_i)^2$ is minimized to create the matrix $M = (I-W)^T(I-W)$. Then we define the kernel matrix $K = \Lambda_{max}I-M$, where Λ_{max} is the maximum eigenvalue of M.

Suppose we have an input data set $X = \{\boldsymbol{x_i} : i = 1,\ldots,n\}$ where each observation x_i is a p-dimensional vector. Kernel LLE algorithm [10, 11] consists of two main steps: the first step is to linearize the distribution of the input data by using a non-linear mapping Φ from the input space $\mathbb{R}^p$ to a higher-dimensional (possibly infinite-dimensional) feature space $\mathcal{F}$. The mapping Φ is defined implicitly, by specifying the form of the dot product in the feature space. In other words, given any pair of mapped data points, the dot product is defined in terms of a kernel function (13).

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad (13)$$

In the second step, eigen-analysis is applied to the mapped data set $\Phi = \{\Phi_i : I = 1,\ldots,n\}$ in the feature space, where $\Phi_i = \Phi(x_i)$. The second step of kernel LLE is to apply eigen-analysis in the feature space by performing an eigen-decomposition on the covariance matrix of the mapped data which is given by (14), where (15) s the centered mapped data.

$$C = \frac{1}{n-1} \sum_{i=1}^{n} \tilde{\Phi}(x_i)^T \tilde{\Phi}(x_i) \quad (14)$$

$$\tilde{\Phi}(x_i) = \Phi(x_i) - (1/n) \sum_{i=1}^{n} \Phi(x_i) \quad (15)$$

The eigenvectors of C are given by (16).

$$v = \frac{1}{\mu} Cv = \sum_{i=1}^{n} \tilde{\Phi}(x_i)\left(\frac{1}{\mu(n-1)} \tilde{\Phi}(x_i)^T v\right) = \sum_{i=1}^{n} a_i \tilde{\Phi}(x_i) \quad (16)$$

where

$$a_i = \left((\tilde{\Phi}(x_i)^T v\right)/(\mu(n-1)) . \quad (17)$$

In other words, an eigenvector of C is a linear combination of $\{\tilde{\Phi}(x_i)\}$. Taking the dot product of $\tilde{\Phi}(x_j)$ with v yields (18).

$$\tilde{\Phi}(x_j) \cdot v = \sum_{i=1}^{n} a_i \tilde{\Phi}(x_i) \cdot \tilde{\Phi}(x_j) = \sum_{i=1}^{n} a_i \tilde{K}_{ij} \quad (18)$$

which implies (19).

$$\mu(n-1) a_j = \sum_{i=1}^{n} a_i \tilde{K}_{ij} \quad (19)$$

Hence,

$$\tilde{K} a = \tilde{\mu} a \quad (20)$$

where $a = (a_1, ..., a_n)$ and $\tilde{\mu} = \mu(n-1)$. That is, a is an eigenvector of $\tilde{K}$. If the eigenvectors of C are orthonormal (i.e. $v^T v = 1$) then (21) and hence (22), hold true.

$$1 = v^T v = \sum_{i,j=1}^{n} a_i a_j \tilde{\Phi}(x_i) \cdot \tilde{\Phi}(x_j) = \sum_{i,j=1}^{n} a_i a_j \tilde{K}_{ij} \quad (21)$$

$$\|a\| = 1/\mu(n-1) \quad (22)$$

The main algorithmic step of our proposed kernel LLE chart as shown in Table I. The kernel LLE algorithm is based on the concepts of LLE and kernel PCA.

TABLE I
ALGORITHMIC STEPS OF THE PROPOSED APPROACH

1) Construct a sparse matrix of local predictive weights $W_{i,j}$, such that $\sum_j W_{i,j} = 1, W_{i,j} = 0$ if x_j is not a *k*-nearest neighbor of x_i and $\sum_j (W_{i,j} x_j - x_i)^2$ is minimized.
2) Construct the LLE matrix, $M = (I - W)^T (I - W)$.
3) Construct the kernel matrix, $K = \lambda_{\max} I - M$.
4) $K = (K_{ij})$ of the mapped data: $K_{ij} = K(x_i x_j) = \Phi(x_i) \cdot \Phi(x_j)$.
5) Construct the kernel matrix $\tilde{K} = HKH$ of the centered mapped data, where $H = I - J/n$ the centering matrix is defined in terms of the identity matrix *I* and the matrix of all ones *J*.
6) Find the largest *p* eigenvectors $a_r (r = 1, ..., p)$ of $\tilde{K}$ and their corresponding eigenvalues $\tilde{\mu}_r$.
7) Given a test point *x* with image $\Phi(x)$, compute the projections onto eigenvectors v_r given by the equation

$$v_r \cdot \tilde{\Phi}(x) = \frac{1}{\sqrt{(n-1)}} \sum_{i=1}^{n} a_i \tilde{\Phi}(x_i) \cdot \tilde{\Phi}(x_j) .$$

Assuming we want $\pm 3\sigma$ confidence intervals, the upper control limit (UCL), the center line (CL), and the lower control limit (LCL) of the kernel LLE chart are defined in (23) to (25).

$$UCL = +3\sqrt{\tilde{\mu}_r} \quad (23)$$

$$CL = 0 \quad (24)$$

$$LCL = -3\sqrt{\tilde{\mu}_r} \quad (25)$$

IV. EXPERIMENTAL RESULTS

We conducted experiments on three different data sets with known outliers. In all the experiments, the number of nearest neighbors was set to one less than the rank of the input matrix and the dimension of the output matrix was set to the number of input vectors.

A. Experiment #1: Woodmod Dataset

We tested the performance of our proposed technique on a data set $X=[x_1, x_2, ...x_{20}]^T$ (called woodmod data [8] which contains 20 observations as shown in Table II.

Each observation $\boldsymbol{x}_i=[\, x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}]$ has 5 variables which correspond respectively to:

- No. of fibers per square millimeter in Springwood
- No. of fibers per square millimeter in Summerwood
- Fraction of Springwood
- Fraction of light absorption by Springwood
- Fraction of light absorption by Summerwood

TABLE II
WOODMOD DATASET

x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}
0.5730	0.1059	0.4650	0.5380	0.8410
0.6510	0.1356	0.5270	0.5450	0.8870
0.6060	0.1273	0.4940	0.5210	0.9200
0.4370	0.1591	0.4460	0.4230	0.9920
0.5470	0.1135	0.5310	0.5190	0.9150
0.4440	0.1628	0.4290	0.4110	0.9840
0.4890	0.1231	0.5620	0.4550	0.8240
0.4130	0.1673	0.4180	0.4300	0.9780
0.5360	0.1182	0.5920	0.4640	0.8540
0.6850	0.1564	0.6310	0.5640	0.9140
0.6640	0.1588	0.5060	0.4810	0.8670
0.7030	0.1335	0.5190	0.4840	0.8120
0.6530	0.1395	0.6250	0.5190	0.8920
0.5860	0.1114	0.5050	0.5650	0.8890
0.5340	0.1143	0.5210	0.5700	0.8890
0.5230	0.1320	0.5050	0.6120	0.9190
0.5800	0.1249	0.5460	0.6080	0.9540
0.4480	0.1028	0.5220	0.5340	0.9180
0.4170	0.1687	0.4050	0.4150	0.9810
0.5280	0.1057	0.4240	0.5660	0.9090

The woodmod data variables are highly correlated as shown in Fig. 1, and hence multidimensional quality control charts should be applied.

Fig. 2 shows that the T^2 control chart is unable to detect outliers. Also, the principal component control chart is unable to detect outliers as depicted in Fig. 3. We can clearly see in Fig. 3 that the observations 4, 6, 8 and 19 have higher variations than the rest of the observations although they still lie within the upper and lower control limits.

The kernel LLE chart is able to detect the observations 2, 4, 6, 8, 10, 11, 12, 13 and 19 as outliers as shown in Fig. 4.

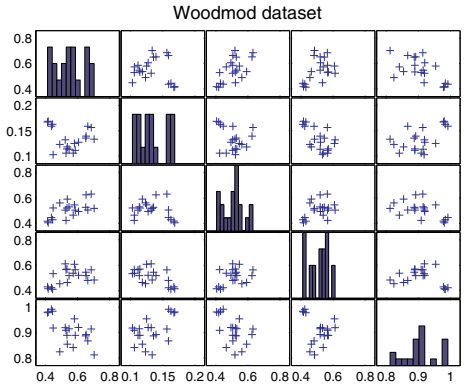


Fig. 1. Scatter Plot of the woodmod Dataset. The woodmod data variables are highly correlated and hence multidimensional quality control charts should be applied.

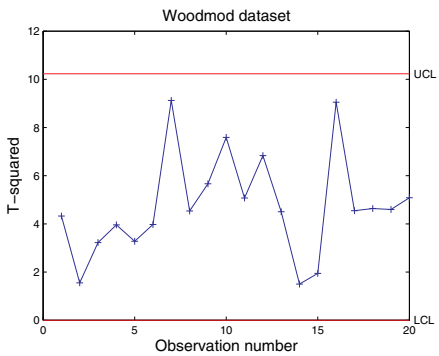


Fig. 2. T^2 Control Chart. There are no outliers detected by this control chart.

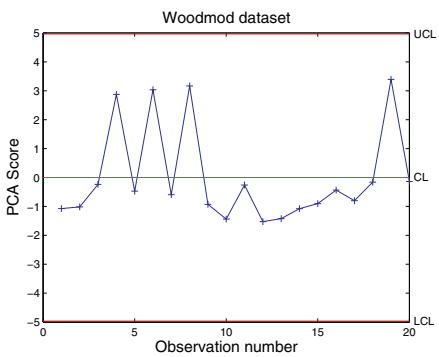


Fig. 3. Principal Component Chart. There are no outliers detected by this control chart, either.

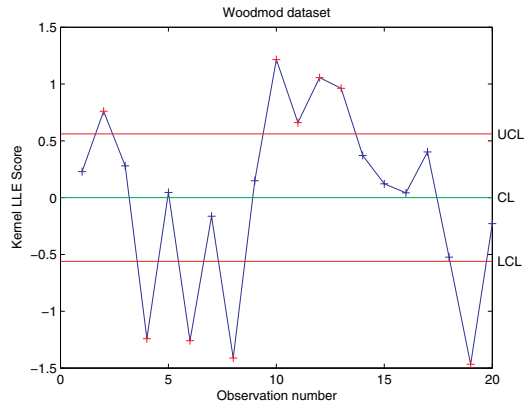


Fig. 4. Kernel LLE Chart. The kernel LLE chart is able to detect the observations 2, 4, 6, 8, 10, 11, 12, 13 and 19 as outliers.

B. Experiment #2: Stackloss Dataset

Our second analysis was performed on a dataset called Stackloss shown in Table III. This dataset describes the plant oxidation of ammonia to nitric acid, and contains 21 observations, where each observation has 4 variables: rate, temperature, acid concentration, and stackloss.

TABLE III
STACKLOSS DATASET

x_{i1}	x_{i2}	x_{i3}	x_{i4}
80.0	27.0	89.0	42.0
80.0	27.0	88.0	37.0
75.0	25.0	90.0	37.0
62.0	24.0	87.0	28.0
62.0	22.0	87.0	18.0
62.0	23.0	87.0	18.0
62.0	24.0	93.0	19.0
62.0	24.0	93.0	20.0
58.0	23.0	87.0	15.0
58.0	18.0	80.0	14.0
58.0	18.0	89.0	14.0
58.0	17.0	88.0	13.0
58.0	18.0	82.0	11.0
58.0	19.0	93.0	12.0
50.0	18.0	89.0	8.0
50.0	18.0	86.0	7.0
50.0	19.0	72.0	8.0
50.0	19.0	79.0	8.0
50.0	20.0	80.0	9.0
56.0	20.0	82.0	15.0
70.0	20.0	91.0	15.0

The scatter plot shown in Fig. 5 confirms the existence of a high correlation between the variables. The T^2 control chart displayed in Fig. 6 was able to identify the last observation m=21 as an outlier. The principal component chart, however, did not detect any outliers as shown in Fig. 7.

On the other hand, the kernel LLE chart (see Fig. 8) was able to identify observations 1, 2, 3, 15, 16, 17, 18, and 19 as outliers.

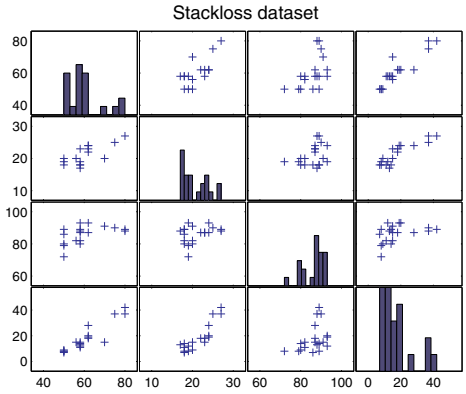


Fig. 5. Scatter Plot of Stackloss Dataset. The scatter plot confirms the existence of a high correlation between the variables.

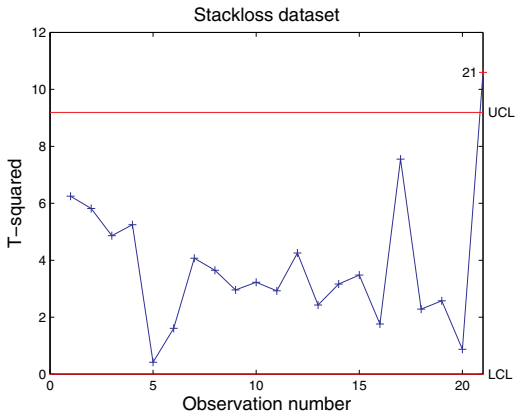


Fig. 6. T^2 Control Chart. There are no outliers detected by this control chart.

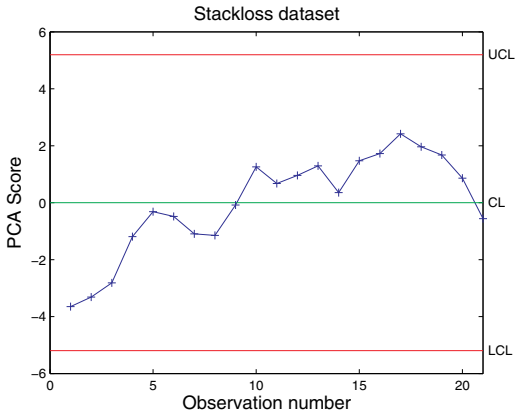


Fig. 7. Principal Component Chart. There are no outliers detected by this control chart.

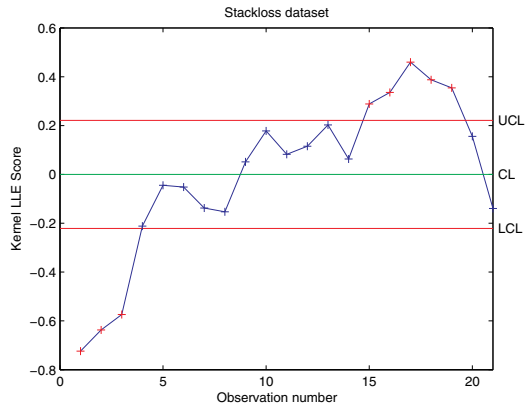


Fig. 8. Kernel LLE Chart. The kernel LLE chart (see Fig. 8) was able to identify observations 1, 2, 3, 15, 16, 17, 18, and 19 as outliers.

C. *Experiment #3: Phosphorus Content Data*

Our third analysis was performed on a dataset (Table IV) describing the organic and inorganic phosphorus content of the soil in comparison with the corn grown. Eighteen observations were selected where each observation has three variables: inorganic phosphorus, organic phosphorus, and plant phosphorus.

TABLE IV
PHOSPHOROUS CONTENT DATA SET

x_{i1}	x_{i2}	x_{i3}
0.40	53.00	64.00
0.40	23.00	60.00
3.10	19.00	71.00
0.60	34.00	61.00
4.70	24.00	54.00
1.70	65.00	77.00
9.40	44.00	81.00
10.10	31.00	93.00
11.60	29.00	93.00
12.60	58.00	51.00
10.90	37.00	76.00
23.10	46.00	96.00
23.10	50.00	77.00
21.60	44.00	93.00
23.10	56.00	95.00
1.90	36.00	54.00
26.80	58.00	168.00
29.90	51.00	99.00

The scatter plot of the data set is displayed in Fig. 9. As shown in Fig. 10, the T^2 control chart was able to identify the observation 17 as an outlier, whereas principal component chart did not identify any outliers as illustrated in Fig. 11.

Kernel LLE chart was, however, able to detect the observations 2, 3, 4, 5, 12, 15, 16, 17, and 18 as outliers as shown in Fig. 12.

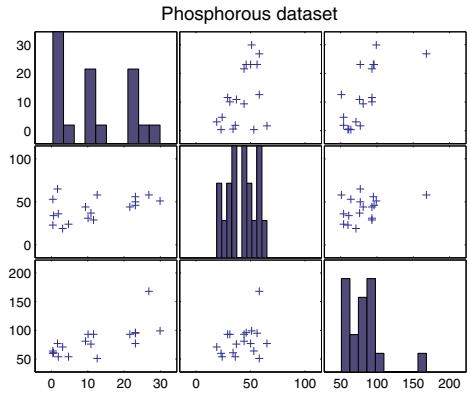


Fig. 9. Scatter Plot of Phosphorous Dataset. The scatter plot confirms the existence of a high correlation between the variables.

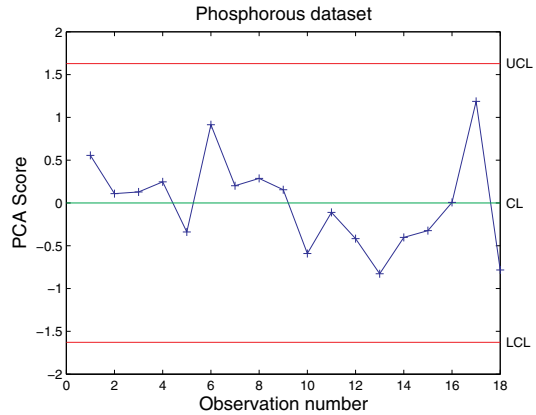


Fig. 11. Principal Component Chart. There are no outliers detected by this control chart.

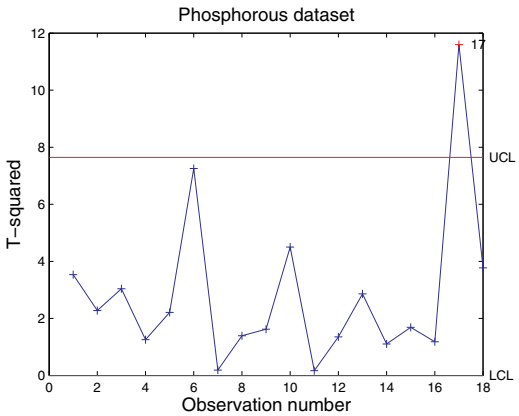


Fig. 10. T^2 Control Chart. There is a single outlier detected by this control chart (observation 17).

CONCLUSIONS

In this paper, we proposed a robust multivariate control chart for outlier detection using kernel locally linear embedding algorithm. The core idea behind our proposed technique is to project the data into a Hilbert space in order to extract the eigenvalues and eigenvectors of a kernel matrix. The experimental results clearly show a much improved performance of the proposed approach in comparison with the current multivariate control charts.

REFERENCES

[1] D.C. Montgomery, *Introduction to Statistical Quality Control*, John Wiley & Sons, 2005.

[2] K. Yang and J. Trewn, *Multivariate Statistical Methods in Quality Management*, Mc Graw Hill Professional, 2004.

[3] K.H. Chen, D.S. Boning, and R.E. Welch, "Multivariate statistical process control and signature analysis using eigenfactor detection methods,"*Proc. Symposium on the Interface of Computer Science and Statistics*, Costa Mesa, CA, 2001.

[4] I.T. Jolliffe, *Principal Component Analysis*, New York: Springer, 1986.

[5] J.A. Vargas, "Robust estimation in multivariate control charts for individual observations," *Journal of Quality Technology*, vol. 35, no. 4, pp. 367-376, 2003.

[6] N.D. Tracy, J.C. Young, and R.L. Mason, "Multivariate quality control charts for individual observations," *Journal of Quality Technology,* vol. 24, no. 22, pp. 88-95, 1992.

[7] J.H. Sullivan and W.H. Woodall, "A comparison of multivariate control charts for individual observations," *Journal of Quality Technology*, vol. 28, no. 24, pp. 398-408, 1996.

[8] F.A. Alqallaf, K.P. Konis, and R.D. Martin, and R.H. Zamar, "Scalable robust covariance and correlation estimates for data mining," *Proc. ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 14-23, 2002.

[9] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, 2nd edition, 1998.

[10] B. Scholkopf, A. Smola, and K-R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp. 1299-1319, 1998.

[11] J. Shawe-Taylor and C. Williams, "The stability of kernel principal components analysis and its relation to the process eigenspectrum.," *Advances in neural information processing systems*, vol. 15, 2003.

[12] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323-2326, 2000.

A New Method for Synchronization and Control of the Chen Chaotic System

Afshin Izadian, Boyd Edwards, and Parviz Famouri, *Senior Member, IEEE*

Abstract— **chaotic behavior of dynamical systems can be utilized in many applications such as secure communication systems, heart arrhythmias, and chemical reactors. In this regard, they are required to be regulated and synchronized with desired references and obtain specific functionality. This paper introduces the application of an input-output controller to synchronize the Chen chaotic system with desired references. Adaptive techniques are applied for gain adaptation and the system is controlled to track arbitrary trajectories. The performance of the control is compared with other types of controllers.**

I. INTRODUCTION

DYNAMICAL systems under specific conditions show chaotic behavior which often possess a wideband power spectrum and random characteristics in time domain [1]. Chaotic systems are very sensitive to the initial conditions such that a very small perturbation results in a large variation in their behavior.

Chaotic behavior of dynamical systems can be utilized in many applications if they are required to follow a desired pattern such as secure communication systems where a random modulation is required and must be synchronized at both ends of the communication transmission line [2]. Control and synchronizing techniques are required to provide a fast and accurate trajectory controls for these types of applications. Chaotic system mainly possess high order (>2) nonlinear dynamics which can better be controlled by nonlinear control approaches. The performance of the control depends on the power of the controller and its ability to follow the sudden variations. A suitable controller ideally contains a simple structure and requires minimum information about the system under control. It generates minimum control command and controls the plant in a short time. It should be robust against the plants parameter variations. Several techniques are recommended for control of chaotic systems such as stochastic control, Lyapunov methods, robust feedback control, and feedback linearization, feedback control of bifurcation, variable structure control techniques [3] and passive control techniques [4]. These techniques mainly depend on the system parameters and state variables which require accurate modeling and state observers.

This paper utilizes a new input-output controller with a new adaptation technique and a model to synchronize the plant's output (chaotic system) with that of the model or an arbitrary reference. Chen chaotic system with almost the same structure as Lorenz attractor (dual of the Lorenz [5]) is studied and used for synchronization technique. The controller is applied to control the output of the reference with that of the plant excited at different initial conditions.

The paper is organized as follows: the next section is an introduction to Chen chaotic system, and in section 3 the controller structure is illustrated. Simulation results and tracking performance are discussed in section 4.

II. CHEN CHAOTIC SYSTEM

The dynamics of the Chen chaotic system is described as follows [5]:

$$\begin{cases} \dot{x} = a(y - x) \\ \dot{y} = (c - a)x - xz + cy \\ \dot{z} = xy - bz \end{cases} \tag{1}$$

Chaotic behavior occurs when a=35, b=3, and c=28. The structure of the Chen chaotic system is almost similar to that of the Lorenz attractor where the chaotic behavior occurs at a=28, b=8/3 and c=10.

The Chen dynamical system shows three sets of fixed points which result in saddle node and unstable spirals described as follows: The fixed point at the origin results in the eigenvalues (-0.8359), (23.8359), and (-3), implying a saddle node. Two other fixed points share the eigenvalues at (-18.4280), and (4.2140 ±14.8846i) since the real parts of these eigenvalues are positive; the linearization theory predicts these fixed points to be unstable spirals. The system's chaotic behavior at the unit initial condition is shown in Figure 1. For control purposes, the original nonlinear system is being controlled by applying a control command $u(t)$ to the z-axis of the system as follows [4]:

$$\begin{cases} \dot{x} = a(y - x) \\ \dot{y} = (c - a)x - xz + cy \\ \dot{z} = xy - bz + u \end{cases} \tag{2}$$

The system's chaotic behavior is shown in Figures 2 and 3.

Manuscript received October 5, 2007. This work was sponsored in part by NASA and NSF.

A. I. is a PhD candidate at the Department of Computer Science and Electrical Engineering, West Virginia University, (e-mail: Izadian@ieee.org).

B. E. is with the Department of Physics, West Virginia University, (e-mail: boyd.edwards@mail.wvu.edu).

P. F. is with the Department of Computer Science and Electrical Engineering, West Virginia University, (e-mail: Parviz.famouri@mail.wvu.edu).

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 7–10.

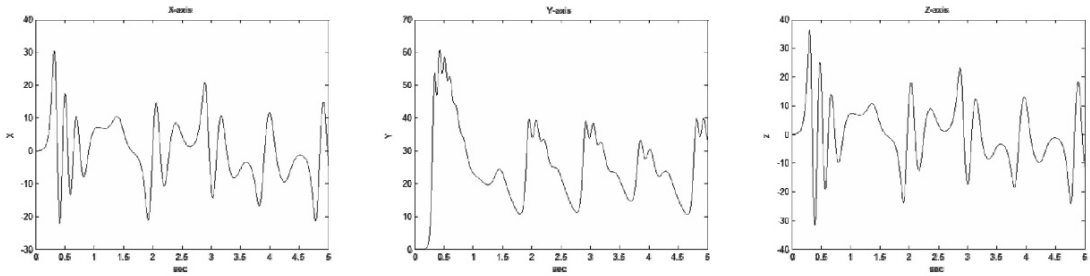


Fig. 1. Chaotic behavior of the Chen system in x, y, and z-axes.

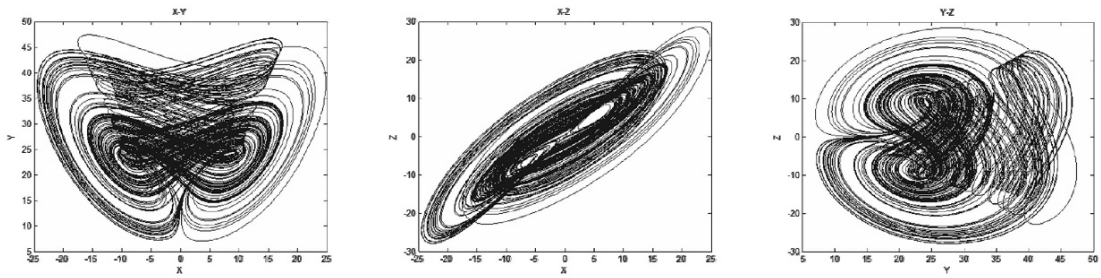


Fig. 2. 2D Chaotic behavior of the Chen system respect to different axes.

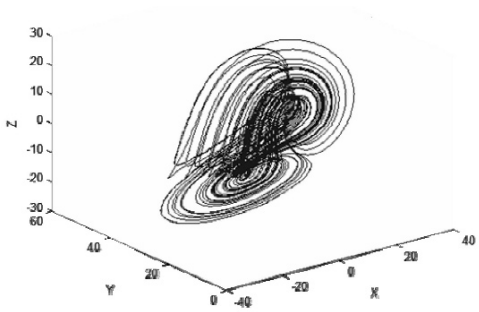


Fig. 3. 3-D chaotic behavior of Chen system.

Definition 1: The Chen chaotic system is minimum phase if it has relative degrees $\{1,1,...\}$.

Definition 2: Two chaotic systems are synchronized if the error dynamic system between their outputs is asymptotically stable [2].

In this paper, we consider one of the chaotic systems as model reference and the other as a plant to be controlled. The control objective is to synchronize these two systems while they start at different initial conditions. The other control objective is to be able to synchronize any state parameters of the chaotic system with an arbitrary reference.

III. Synchronization Technique

The control command to the z-axis of the Chen chaotic system can be defined as follows: [6]

$$u = k_p y_p + k_e (y_p - y_m) + k_r r \tag{7}$$

where k_r, k_e, k_p are the controller coefficients and are being adjusted according to the gain adaptation techniques, and y_m, y_p which are the model reference and plant's output signals. Any of the system state variables can be chosen as an output. r is defined as the model reference input. In a perfect tracking condition, the coefficients of k_r, k_e, k_p are chosen such that the zero error conditions hold. Estimations of these coefficients are introduced as $\hat{k}_r, \hat{k}_e, \hat{k}_p$ and result in the equivalent control command as follows:

$$u_{eq} = \hat{k}_p y_p + \hat{k}_e e + \hat{k}_r r \tag{8}$$

where control gain coefficients are updated according to the following adaptation techniques:

$$\dot{\hat{k}}_r = -P_0 \operatorname{sgn}(s) r \tag{16}$$

$$\dot{\hat{k}}_p = -P_0 \operatorname{sgn}(s) y_p \tag{17}$$

$$\dot{\hat{k}}_e = -P_0 \operatorname{sgn}(s) e \tag{18}$$

with $P_0 > 0$ is a positive value and s is defined as the sliding surface $s = Ge = 0$ in which G is the switching gain matrix and $e = y_p - y_m$ is the tracking error.

Seeking zero error conditions the switching matrix can be considered as identity value. The proof of stability is introduced in [8].

IV. Simulation Results

As mentioned earlier, the chaotic systems were excited initially at different values. First, the systems were allowed to oscillate individually for 0.5 seconds. The output of the chosen axes oscillated differently as shown in Figure 4. Then (at t=0.5) the control command was applied to the plant to force its output track the desired reference signal. The synchronization and tracking profile of the x-axis of the plant with the y-axis of the model is shown in Figure 4. The synchronization is completed in less than one second. Figure 5 shows the tracking error and the adaptation process. At higher adaptation gains, faster adaptation occurs which requires higher control efforts.

For an arbitrary reference signal (not chaotic), the adaptation time is less than 0.2 seconds. Figure 6 shows the tracking profile for an arbitrary reference signal. The control command was applied at time 0.5 second, and before that systems were oscillated differently.

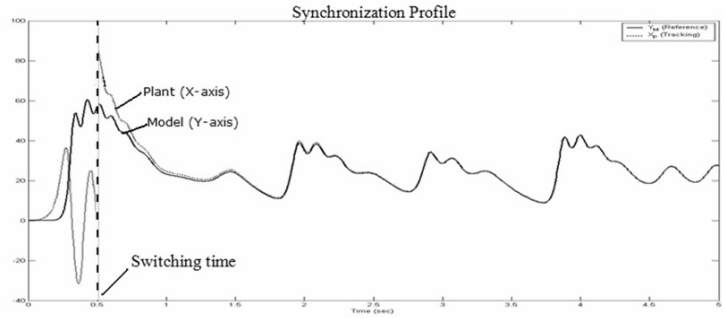


Fig.4. The synchronization is completed in less than a 0.8 sec. The synchronization time is at t=0.5 sec. The controller is turned on and applied to the system to synchronize the Y_d and X_r.

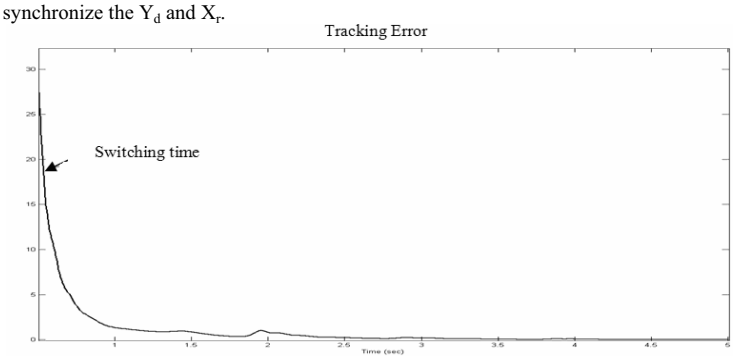


Fig. 5. The synchronization error and its stability. As time increases the synchronization, error reaches zero.

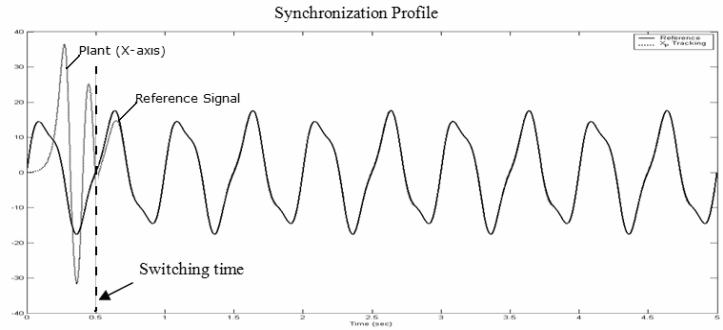


Fig. 6. Synchronization with an arbitrary reference. The controller is switched on at time t=0.5 sec and is forced to follow an arbitrary sine wave. The x-axis output of the response system tracks the variations of the reference and is synchronized with the reference signal in tile less than 0.2 seconds.

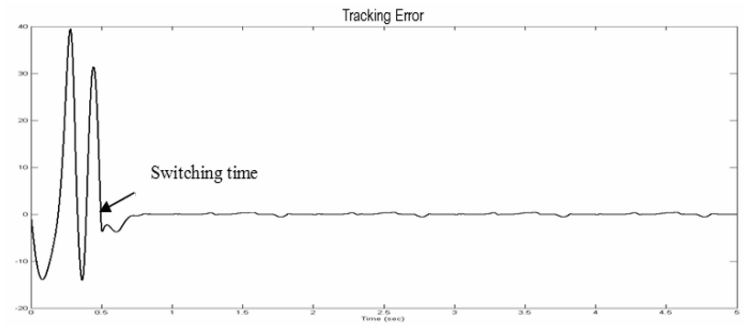


Fig. 7. Synchronization Error of an arbitrary reference. The synchronization time is shown, the error reaches zero in a short time.

Xiaoxin Liao et. al, have reported an error dynamic analysis approach in which the same state variables of the Chen chaotic system have been synchronized [5]. In the new synchronization approach, the synchronization time for x-state variable is much shorter than that of reported in [5]. Moreover, the synchronization of different axes was applicable. Synchronization of Chen chaotic system with an arbitrary reference shows the flexibility of the controller.

CONCLUSION

A new controller-synchronizer was designed for the Chen chaotic system. The controller made use of a new adaptation law for its parameter adjustment. The controller was applied to synchronize the Chen chaotic system with the output of another identical chaotic system and with an arbitrary trajectory. A comparison with other controllers showed a faster and more accurate controller with more degrees of freedom in synchronization.

REFERENCES

[1] G. Kis, Z. Jako, M. P. Kennedy, G. Kolumban," Chaotic Communications without Synchronization," Telecommunications, 1998, 29 March -1 April 1998, Conference publication, No. 451, IEE, 1988.

[2] G. Chen, X. Dong, From Chaos to Order: Methodologies, Perspectives and applications, World Scientific, Singapore 1998.

[3] G. Chen, X. Dong," Control of Chaos- A Survey," Proceeding of the 32nd Conference on Decision and Control, San Antonio, Texas-December 1993. pp 469-474.

[4] D. Qi, G. Zhao and Y. Song, "Passive control of Chen chaotic System," Proceedings of the 5th World Congress on Intelligent Control and Automation, June 15-19, 2004 Hangzhou, P. R. China pp 1284-1286.

[5] X. Liao, G. Chen, H. O. Wang," On Global Synchronization of Chaotic Systems," Proceedings of the American Control conference, Anchorage, AK May 8-10, 2002, Page 2255-2259

[6] K. K. D. Young, "Design of Variable structure Model-Following Control Systems," IEEE Transaction on Automatic Control, AC-23. DEC. 1978, pp 1079-1085.

[7] Y. T. Chan, "Perfect model following with a real model," Proc. Joint Automat. Contr. Conf. 1973, pp. 287-293

[8] A. Izadian and P. Famouri,"On a New Input-Output Lyapunov Design Adaptive Controller," IEEE Trans. On Control System Technology, 2007 Submitted

The Intra Prediction in H.264

Ahmad Khalil Khan and Habibullah Jamal
Department of Electrical Engineering,
University of Engineering and Technology,Taxila
profkhalil@uettaxila.edu.pk , drhjamal@uettaxila.edu.pk

Abstract: A multidirectional spatial prediction method to reduce spatial redundancy by using neighbouring samples as a prediction for the samples in a block of data to be encoded is already included in the H.264/AVC. The spatial prediction may be performed on sample wise DPCM instead of the block-based manner as used in the current H.264/AVC standard. The block structure is retained for the residual difference entropy coding process. There are two approaches of spatial prediction i.e. sample prediction and block prediction. The Second approach is more efficient. It is being introduced into the H.264/AVC standard as an enhancement. This paper is a brief survey of what is important in intra prediction and with some experimental work. The work involves running of the code of H.264.

1. INTRODUCTION

Video coding standard H.264/AVC was developed by Joint Video Team (JVT). It outperforms the previous compression standards in many respects due to the added new features. Any reference may be consulted in this respect but we will restrict our discussion to the intra prediction.
The formation of a prediction block P based on previously encoded and reconstructed blocks and subtracted from the current block prior encoding is called intra mode prediction. P is formed for each 4x4 block or for a 16x16 macroblock. There are nine prediction modes available for 4x4 luma block, four modes for a 16x16 luma block and four modes for the chroma blocks. The prediction mode for each block that minimizes the difference between P and the current block is selected[1].

The difference of intra block output and the current block is the residual and it is going to be transformed and then quantized. This signal is transmitted with some signal processing and demodulated and some previously reconstructed picture is added to it to reconstruct a picture. Similarly when one part is going to the transmitter the other part is inverse quantized and then transformed to get the reference picture which is used to predict macroblock. It takes more than 8% of the computational time [2].

The types of intra coding are supported based on slice-coding types like Intra_4x4 or Intra_16x16 together with chroma prediction.The Intra_4x4 mode means each 4x4 luma block which is used for coding of parts of a picture. The Intra_16x16 mode means 16x16 luma block which is used for coding smooth areas of a picture. A separate chroma prediction of 8x8 is also used.

In H.264/AVC intra prediction is conducted in the spatial domain by referring to neighbouring samples of previously-coded blocks which are to the left and for above the block to be predicted. Because of this a constrained intra mode is transmitted that allows prediction only from intra-coded neighbouring macroblocks.

For Intra_4x4 mode, each 4x4 block is predicted from spatially neighbouring samples as illustrated in Fig.1a and Fig.1b.

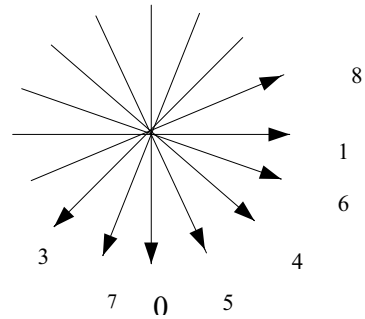


Fig.1a. For Intra_4x4 eight predictions directions

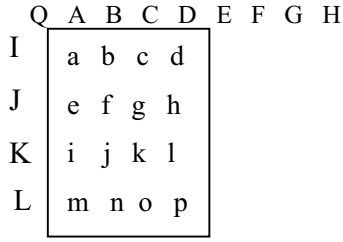


Fig.1b. Intra_4x4 for samples a-p of a block using samples A-Q.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 11–15.

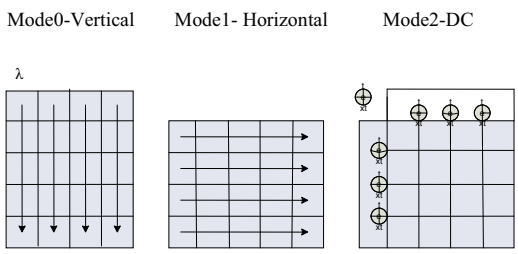


Fig.2. Three of the Nine Intra_4x4 prediction modes.

The 16 samples from a to p of 4x4 block are predicted using prior decoded samples in adjacent blocks labeled as AQ. There are nine prediction modes available. In DC prediction one value is used to predict the entire 4x4 block and other two out of eight modes are illustrated in Fig.2.[3].

In Fig.2 a few of the nine Intra_4x4 prediction modes are shown. For the vertical prediction or mode0, the sample above the 4x4 block are copied into the block as indicated by the arrows. In the horizontal prediction (mode1) the samples to the left of the 4x4 block are entered. For DC prediction or mode2 the adjacent samples average is taken. The other modes not shown are diagonal prediction modes. They are named as diagonal-down-left (mode3), diagonal-down-right (mode4), vertical-right (mode5), horizontal-down(mode6), vertical-left (mode7) and horizontal-up(mode8) [3].

Intra_16x16 has four prediction modes. Vertical prediction (mode0), horizontal prediction (mode1), DC prediction (mode2) and plane prediction (mode3) are supported. They are specified just like the modes in Intra_4x4 prediction. Instead of four neighbours on each side, 16 neighbours on each side to a block are used. Using a similar prediction technique the chroma samples are predicted[1].

Mode decision technique may be based on rate-distortion optimization. For all possible modes rate-distortion cost is calculated and the best one having minimum rate-distortion is chosen. For the purpose high computational complexity is required. To keep the complexity low two methods are used: i) early SKIP mode decision ii) Selective intra mode decision .[4][5].

Without significant rate-distortion performance degradation the complexity of the encoder is discussed in [6]. Intra-frame prediction and I_PCM prediction is treated well in [3]. In the prediction procedure integer motion estimation (IME), fractional motion estimation (FME) and intra prediction (INTRA) are included. The computational complexity of different components may be given as : Integer Motion Estimation is 52.034%, intra prediction 0.544% and sub_pixel residue/Hadamard 37.207%.[7].

In H.264 the rate-distortion (R-D) optimization is the main complexity source.[8]. Architectures and designs for H.264/AVC intra frame coders etc are considered.[7],[9],[10],[11],[12].

A YUV picture of a video sequence is represented by three rectangular arrays of integer valued samples. One of these arrays is called the luma component and shows the brightness at each sample location. The other two arrays are typically called the chroma arrays of the same. At each sample they represent the color-difference chromaticity.For 4:4:4 all the three arrays are of the same dimension. For 4:2:0 each chroma array is half the width and half of the length of the luma array. For 4:2:2 the chroma arrays are half in width but of the same height.[13].

2. THE INTRA PREDICTION MODE SELECTION

The intra mode that meets the best Rate-distortion tradeoff can be determined. The mode is determined by minimizing the Lagrangian functional :

$J(s,c,MODE \mid QP, \lambda_{MODE}) = SAD(s,c,MODE \mid QP) + \lambda_{MODE}.R(s,c,MODE \mid OP)$

Where QP = Quantization parameter

λ_{MODE}= Lagrange multiplier for mode decision

SAD(.) = Sum of absolute differences between the original blocks and its reconstruction c

MODE. = Prediction mode[14].

3. THE LOSSY AND LOSSLESS CODING

The H.264/MPEG-4 AVC video coding standard developed in 2003, was developed jointly as ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 part 10) Advanced Video Coding (AVC) is a remarkably efficient lossy video coding standard. Its lossless encoding capabilities are also important. Pulse-code modulation (PCM) macroblock coding mode represented samples of selected macroblock losslessly. It was inefficient from a compression perspective.

The fidelity range extension (FRExt) included the design improvements for more efficient lossless encoding. Further improvements in these lossless encoding techniques have already been adopted.[13].

Block-based prediction in H.264/AVC which is intra picture prediction consisting of using selectable position-dependent linear combinations of neighbouring sample values to form a prediction block.

The goal of the prediction processing is to reduce the quantity of data needed to adequately represent the corresponding block of input picture samples. Intra prediction is used for random access, video sequence editing, still picture coding and significant changes of scene content.

Lossless compression is included in the Fidelity Range Extension (FRExt) capability.H.264/AVC adopts FRExt for high resolution video to HDTV. Sevral parts are added by FRExt to improve the coding efficiency and visual quality at HDTV level including 8x8 transformed newer quantization technique and 8x8 intra prediction. Intra prediction is used in intra slices and may also be used in P and B slices and further 8x8 intra prediction increases the speed of the operation.[15].

There are three basic types of intra prediction 16x16, 8x8 and 4x4 in H.264/AVC. Nine intra prediction modes for 8x8 and 4x4 intra prediction are used.

The AC energy coefficient tells the complexity of the texture. For a complex texture the size of the block for intra prediction is chosen as smaller otherwise a larger block may be chosen. For the purpose the ratio of AC/DC is used to predict the intra prediction block size.[15]

To compute DC and AC coefficient energy in pixel domain, the following equations are used [15].

$$DC\_energy = 1/8x8 \left(\sum_{m=0}^{7} \sum_{n=0}^{7} a_{mn} \right)$$

$$AC\_energy = \left(\left(\sum_{m=0}^{7} \sum_{n=0}^{7} a_{mn} \right) - DC\_energy \right)$$

Where amn is the data in a block.

A normalized AC_DC ratio is

AC _ DC _ ratio = log (AC _ energy) / log (DC _ energy x 64)

It is normalized between 0-1.

QP is also used to determine the smoothness. The effect of QP to cost function without RDO is linear. The threshold can be a linear equation of QP

$$Th = a*Qp + b$$

$$a = 0.0091 , b = 0.236$$

These values are used extensively to determine the block type.[15]

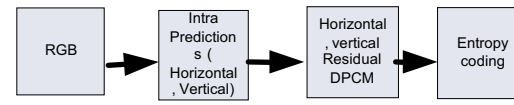


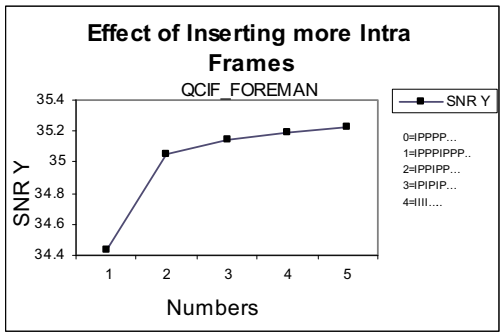
Fig3 Block diagram of the lossless Intra Residual DPCM

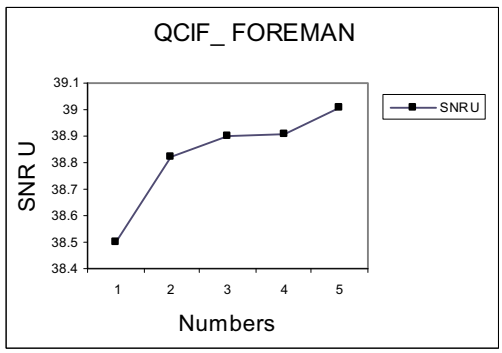
4. EXPERIMENTAL WORK

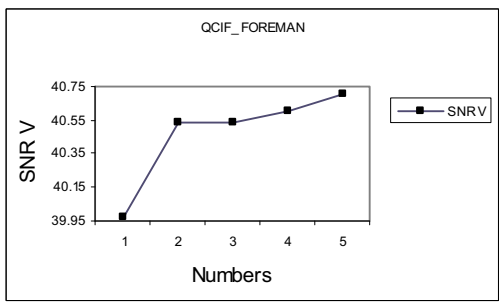
JM9.5 is the standard Code for H.264. It is written in C. It may be run with the required changes in the C code. The code was prepared by the Joint Video Team (JVT) and updated continuously which makes it responsive to day by day additions.[16]. In all coding processes the first frame is always intra coded., other wise it is not possible to start and update the picture transmission with compression. A parameter named as intraperiod in the code determines the number of frames to be coded as intra. When its value is zero then only fist frame is coded as intra. When intraperiod is one then the sequence is like IPPPIPPP.. and likewise we can get other sequences. The number of bits required is increased as the contribution of the intra frames is increased. All this is in line with our perception. Consequently the compression continues to decrease. The SNRs in case of Y,U and V also continue to increase when the contribution of intra increased.

5. CONCLUSION

After a brief survey of different techniques it is obvious that in all types of compression techniques the first frame is intra coded for the reasons already mentioned .The signal to noise ratio continues to increase with intra coding. These are our common observations. Different modes used in case of intra prediction contribute towards the increase of compression.







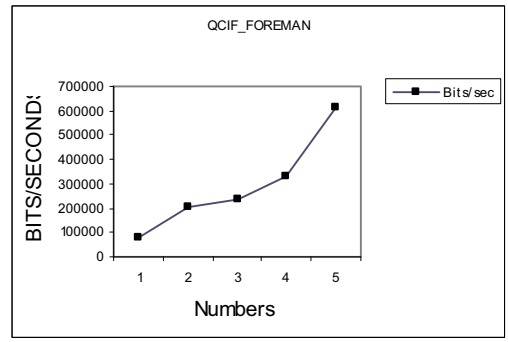


Fig. 4.
Effect of inserting more intra frames by increasing intra period (From top to bottom)
a; Average SNR Y, b; Average SNR U, c; Average

6. REFERENCES

[1]. Iain E. G. Richardson, H.264 and MPEG-4 video compression-video coding for next generation Multimedia, John Wiley, 2003.

[2] Hung – Chih Lin; Yu-Jen Wang; Kai-Ting Cheng; Shang-Yu Yeh; Wei-Nien Chen; Chia-Yang Tsai; Tian-Sheuan Chang; Hseuh-Ming Hang; Algorithms and DSP implementation of H.264/AVC. Design Automation 2006. Asia and South Pacific Conference on 24-27 Jan. 2006 Page(s):8 pp.

[3]. Wiegand, T.; Sullivan, G.J.; Bjntegaard, G.; Luthra, A.; Overview of the H.264/AVC video coding standard
Circuits and Systems for Video Technology, IEEE Transactions on Volume 13, Issue 7, July 2003 Page(s):560 – 576

[4]. Jeyun Lee; Byeungwoo Jeon; Fast mode decision for H.264

Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on Volume 2, 27-30 June 2004 Page(s):1131 - 1134 Vol.2

[5]. Yu-Wen Huang; Bing-Yu Hsieh; Tung-Chien Chen; Liang-Gee Chen Analysis, fast algorithm, and VLSI architecture design for H.264/AVC intra frame coder, Pages 378-401

IEEE Transaction on The circuit and system for video Technology,2004

[6]. Changsung Kim, Hsuan-Huei Shih, C.-C. Jay Kuo, Fast H.264 Intra-prediction mode selection
using joint spatial and transform domain Journal of Visual Image Representation,pp.291-310, Science Direct,17(2006).

[7]. Tung-Chien Chen; Yu-Wen Huang; Liang-Gee Chen;
Analysis and design of macroblock pipelining for H.264/AVC VLSI architecture
Circuits and Systems, 2004. ISCAS '04. Proceedings of the 2004 International Symposium on
Volume 2, 23-26 May 2004 Page(s):II - 273-6 Vol.2

[8]. Arsura, E.; Del Vecchio, L.; Lancini, R.; Nisti, L.;
Fast macroblock intra and inter modes selection for H.264/AVC
Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on
6-8 July 2005 Page(s):4 pp.

[9]. Yu-Wen Huang; Bing-Yu Hsieh; Tung-Chien Chen; Liang-Gee Chen; Hardware architecture design for H.264/AVC intra frame coder
Circuits and Systems, 2004. ISCAS '04. Proceedings of the 2004 International Symposium on
Volume 2, 23-26 May 2004 Page(s):II - 269-72 Vol.2

[10]. Chen-Han Tsai; Yu-Wen Huang; Liang-Gee Chen;
Algorithm and architecture optimization for full-mode encoding of H.264/AVC intra prediction
Circuits and Systems, 2005. 48th Midwest Symposium on
7-10 Aug. 2005 Page(s):47 - 50 Vol. 1

[11]. Yu-Wen Huang; Tung-Chien Chen; Chen-Han Tsai; Ching-Yeh Chen; To-Wei Chen; Chi-Shi Chen; Chun-Fu Shen; Shyh-Yih Ma; Tu-Chih Wang; Bing-Yu Hsieh; Hung-Chi Fang; Liang-Gee Chen;. A 1.3 TOPS H.264/AVC single-chip encoder for HDTV applications
Solid-State Circuits Conference, 2005. Digest of Technical Papers. ISSCC. 2005 IEEE International
6-10 Feb. 2005 Page(s):128 – 588 Vol. 1

[12]. Genhua Jin; Hyuk-Jae Lee;A Parallel and Pipelined Execution of H.264/AVC Intra Prediction.
Volume 2, 24-26 May 2005 Page(s):740 - 744 Vol. 2
Computer and Information Technology, 2006. CIT '06. The Sixth IEEE International Conference on
Sept. 2006 Page(s):246 - 246

[13]. Yung-Lyul Lee; Ki-Hun Han; Sullivan, G.J.;
Improved lossless intra coding for H.264/MPEG-4 AVC
Image Processing, IEEE Transactions on
Volume 15, Issue 9, Sept. 2006 Page(s):2610 - 2615

[14]. Min-Cheol Hwang; Jun-Ki Cho; Jin-Sam Kim; Jin-Hyung Kim; Sung-Jea Ko; Fast Intra Prediction Mode Selection Scheme Using Temporal Correlation in H.264
TENCON 2005 2005 IEEE Region 10
Nov. 2005 Page(s):1 - 5

[15]. Yu-Kun Lin; Tian-Sheuan Chang; Fast block type decision algorithm for intra prediction in H.264 FRext
Image Processing, 2005. ICIP 2005. IEEE International Conference on
Volume 1, 11-14 Sept. 2005 Page(s):I - 585-8

[16].Joint Video Team (JVT), reference Software JM9.5, http
://bs.hhi.de/~suehring/tml/download/Jm9.5.zip

Design and Implementation of Wireless Sensor Network Using Virtual Instruments and ZigBee Communication Protocol

A. Montoya, D. Aristizábal, J. Atencio, C. Correa, N. Montoya
Scientific and Industrial Instrumentation Group
Physics Department
National University of Colombia Sede Medellín
Autopista Norte, Medellín, Colombia
A.A. 3840
amontoya@unalmed.edu.co

ABSTRACT – In this dissertation, a wireless network of sensors was implemented for the monitoring of physical variables, consisting of a software-hardware platform in keeping with the design model of virtual instruments. The communication protocols used were RS232 and ZigBee; the network is made up of various modules, each one consisting of five sensors and one Zigbee transmitter. These modules, in turn, communicate wirelessly with a ZigBee receiver module that is linked up to the RS232 port of a personal computer that administers the entire system. In developing the software, we used a object-oriented programming paradigm, Java programming language, the LINUX operating system, all of which are open source; this, together with the electronic elements used, makes for a low-cost system that can be used for industrial purposes by both small and medium-sized enterprises (SMEs) enhancing both their productivity as well as their competitiveness.

KEYWORDS: Wireless Network, Virtual Instrumentation, Sensors, Monitoring, RS232, ZigBee.

I . INTRODUCTION

Wireless networks are constantly evolving and are now gradually replacing cable network technology given their inherent flexibility and the fact that they can be easily installed. Their development and technology have been such, that little devices called sensors have been incorporated in these networks. This has led us to become acquainted with this recent technology, that is innovative and able to provide quick and reliable solutions to the problems that arise. Their capacity to monitor and manipulate physical elements or phenomena give rise to enormous possibilities for almost any scientific discipline. In particular, our investigation group has directed its applications towards environmental monitoring and precision agriculture; and in the latter case with a view to contributing to the development of Colombia, which represents significant potential in this sector.

The hardware-software platform was designed associating the sensors to virtual instruments, whose communication (data transfer) system implements the ZigBee protocol. Virtual Instrumentation (VI) is a concept that was introduced by National Instruments through its software laboratory Virtual Instrument Engineering Workbench (LabVIEW); VI allows a personal computer (PC) to be used as a measuring instrument. In this way the concept of VI has been conceived as "an instrument that is not real, that is implemented by means of a computer and whose functions are defined by the software" [2], [3], [4].

The platform thus developed can be adapted by making minor modifications to other applications such as the distributed automation [5], [6], [7], [8].

II. MATERIALS AND METHODS

A. The software

In designing and implementing the VI the object-oriented programming paradigm was used. The programming language used is Java and for the IDE (Integrated Development Environment) the NetBeans 5.5, was used both of which are sourced by Sun Microsystems All this is open source software [9].

The dynamic polymorphism, implemented through inheritance is applied to the design of the VI allowing the instruments to be extended in a transparent fashion with an efficient reuse of the code. The classes responsible for communicating are totally detached from the instrument code, which facilitates the adaptation of any new communication protocol. The VI was built in compliance with the requirements of the Java language, so that they could become JavaBean components, thereby accelerating the development of control panels through a IDE.

B. The hardware

The sensor/transmitter unit

This is responsible for obtaining data from the sensor and sending this to the receiver unit. Is consists mainly of: a PIC16f877a microcontroller, a TTL to CMOS level shifter and a XBee-PRO transmitter unit.

The receiver unit

This is responsible for receiving the data sent by the sensor unit and sending it onto the PC through its serial port. It consists

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 16–21.

mainly of: a XBee-PRO receiver module, a TTL to CMOS level shifter and a Max232 integrated circuit. It is interesting to see the properties of the XBee-PRO chip (Fig. 1) which is the basic element used for the network's wireless communication and implements the ZigBee protocol. It is manufactured by MaxStream and has the following fundamental specifications (Table 1).

Fig. 1: XBee-PRO

Table 1

Frequency	2.4 GHz
Power	60 mW (18 dBm)
Indoor	100 m
Outdoor	1.6 Km
Transmission rate	250 Kps
Number of channels	12
Size	2.438 cm x 3.294 cm
Network Topology	Peer-to-peer
	Point-to-point
	Point-to-multipoint
Voltage	2.8 - 3.4 Volts

The sensors

In order to test the network the following was used: 1 temperature sensor (LM335AZ from National Semiconductors), 2 brightness sensors (VT43N1 and VT400 from PerkinElmer Optoelectronics), 1 infrared (OP593 from Optek technology Inc) and 1 magnetic field detector (AU-MS-03R from AUPAX INDUSTRIAL(HK) CO., LIMITED).

III. RESULTS

A. The Software

The virtual instrument

A JavaBeans was developed called *GaugeCircular.java* as an extension of the *JComponent* and that implements the *Serializable* interface that provides the instrument with *persistence* It possesses the *set* and *get* public methods that allow for its visualization (colors) to be varied and the measurement scale to be defined (minimum value, maximum value, number of divisions and subdivisions and the measurement unit), and the *setLectura (double dato)* public method which receives the result of the measurement and calls the instrument to repaint. Fig. 2 contains an illustration of the graphic interface corresponding to the monitoring of a sensor unit: here five examples of this virtual instrument are found in the upper panel.

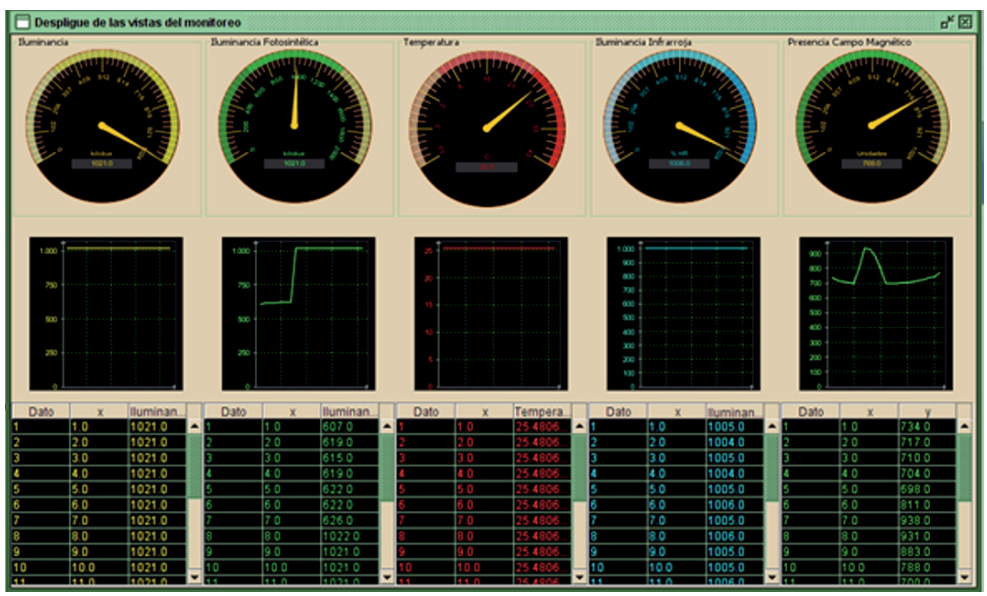


Fig. 2: **G**raphical interfase of one unit sensor.

The Drawing javabean

This JavaBean called the *GraficaXY.java*, is able to display in real time the graphic result of monitoring the measured variable; this was developed using the open source software package *JFreeChart* version *jfreechart-1.0.0-rc1*[10]. As a class, it is an extension of the *JComponent* and also implements the *Serializable* interface. It possesses the *set* and *get* public methods that allow for its appearance to be modified (colors, headings and subheadings of the graph and its axes) the *setDato (double x , double y)* public method in order to receive the data, the *borrarDatos()* public method to erase data together with the private internal auto-scaling methods Fig. 2 illustrates five examples of this bean in the middle panel.

The javabean that displays data in table form

This JavaBean called Tabla.java, is responsible for displaying in real time the results of the measurement. The measurement's number time and value. This is an extension of JPanel and implements the Serializable interface. It possesses the *set* and *get* public methods that allows for its appearance to be modified (colors, headings of the columns. It also has the *setValoresTabla (double x, double y)* public method to receive data and the *borrarDatos()* public method to erase the data. Fig. 2 illustrates five examples of this bean in the middle panel.

The class responsible for the data

Since we had in mind a network made up of stations containing sensor units each containing sensors, a data storage structure (current data) was adopted, consisting of a matrix order arrangement $i \times j \times k$, whereby i stands for the number of the station, j for the number of the sensor unit and k the number or the sensor. Please refer to Fig. 3.

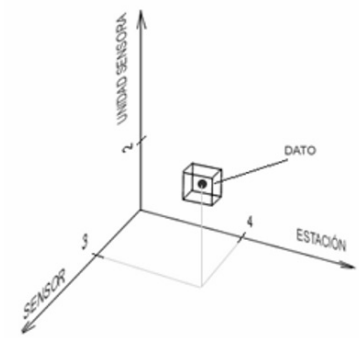


Fig. 3: Current location data of sensor 3 of the unit 2 and Station 4

The class responsible for the treatment of the data is called RAM Data Warehouse. Java (AlmacenRAMDatos.java), which receive the data trough the method setDataSensorsUnitSensored (setDatoSensoresUnidadSensora) (int estacion, int unidad, int sensor, double dato), that request the private method that manages the sensor calibration (process of converting the digital data into real values according to the measure magnitude and technical sheet provided by the manufacture of the sensors or, instead of this, by comparison of the results with a pattern instrument) and saves the obtained values in the public and static variable *dato_actual_sensores[ ][ ][ ]*. This variable allows the access to the measurement's current value of any sensor on the network.

The Classes responsible of communication

The communication with the PC is done through the serial port (RS232 protocol). There were implemented 3 classes for this: ComunicacionPuertoSerial, ConfiguracionRS232 and ParametrosPuertoSerial. The last two ones are in charge of the port configuration and location. The first one handles the serial communication; it implements the SerialPortEventListener, Runnable interfaces and reads the port every 100 ms (this sampling time can be changed), and obtains data from the receiver ZigBee module in ASCII strings that obey a defined protocol and envelope the information about station number, sensor unit number and measurement values of all sensors of the unit; this class is also responsible of the conversion of the information into four data: int estacion, int unidad, int sensor, double dato, and the delivery to Almacen_RAM_Datos class.

The graphical interface

Fig. 4 illustrates the displayed graphical interface when the application is executed. In the menu bar are four buttons: the first one (from left to right) is to access to the serial port configuration; the second one starts the communication between the application and the receiver ZigBee module; the third one ends the communication and the fourth one allows the exit of the application.

Fig. 4: Initial graphical interfase

Two panels were implemented; each one with 12 buttons for the access to the stations and its sensor units: in the figure, sensors of unit 1 and station 1 are being accessed.

When communication starts, a JFrame is displayed; it contains (in the used example): five instances of GaugeCircular bean, five GraficaXY bean instances and five instances of Tabla bean, Fig. 2. They display data of the following sensors: two of illumination (from different precision), one of temperature, one of infrared radiation and another of magnetic field presence. In the second sensor (from left to right) is detected a lighting disturbance; in the fifth sensor, a magnet was approached and immediately moved away.

Fig. 5 illustrates one of the developed sensor units. There are connected five sensors, although the prototype has capacity for eight. The power supply can be done through the normal network or through two AA batteries whose duration will depend essentially on the sampling time; if the data is acquired every 10 minutes, the batteries will last 15 days.

There can be clearly observed the basic components in the figure, among them is necessary to emphasize the XBe-Pro which excels its antenna.

Receiver Unit

It's the unit that receives data sent from the sensor units. It has 12 channels, and can be programmed, for example, for receiving in each channel data from different sensor units that belong to the same station. In the designed architecture, each unit sensor station corresponds to one receiver unit. In the present project, only one station with two sensor units could be installed. Fig. 6 illustrates one receiver unit connected to a PC through its serial port. The power supply can be done through the normal network, through two AA batteries or through the PC if it's connected to the serial port.

Fig. 5: Picture of a sensor unit

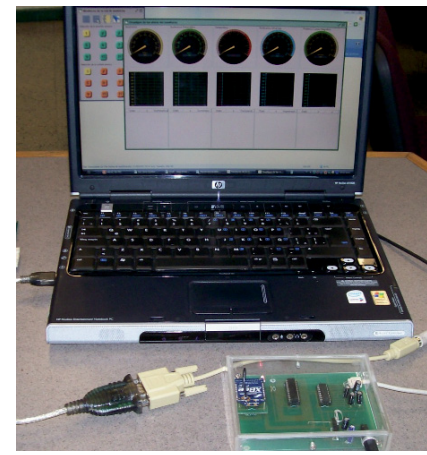


Fig. 6: Picture of the receiving unit connected to the PC

C. Network Architecture

Fig. 7 illustrates the developed sensor network architecture.

This figure corresponds to a station with sensor units (there are represented 12 sensor units) each one with five sensors and in direct communication with the receiver unit, which is connected to a PC through the serial port.

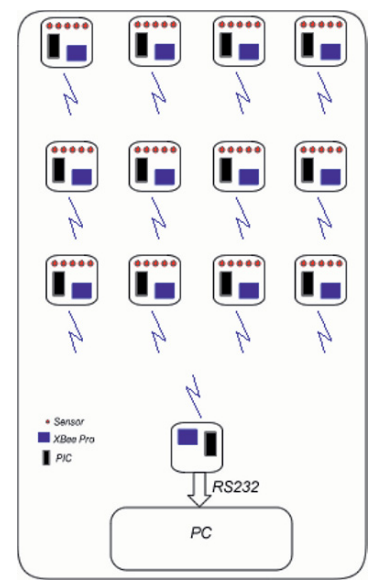


Fig. 7: Topology of a station.

MONTOYA A. ET AL.

D. Description of the Basic system´s operation

Fig. 8 illustrates a block diagram with the basic operation of the system. The microcontroller takes data from sensors and communicates with XBee-PRO via RS232; as the levels of XBee-PRO module are CMOS and the microcontroller has TTL levels as output, the conversion is done through the PIC 74LS245N. The receiver XBEE-PRO module's signal has to be adapted to TTL levels too, this is achieved using the same PIC 74LS245N; in order to make communication with the PC is used Max232 which changes the TTL levels of the converser module to the standards of the RS232 protocol.

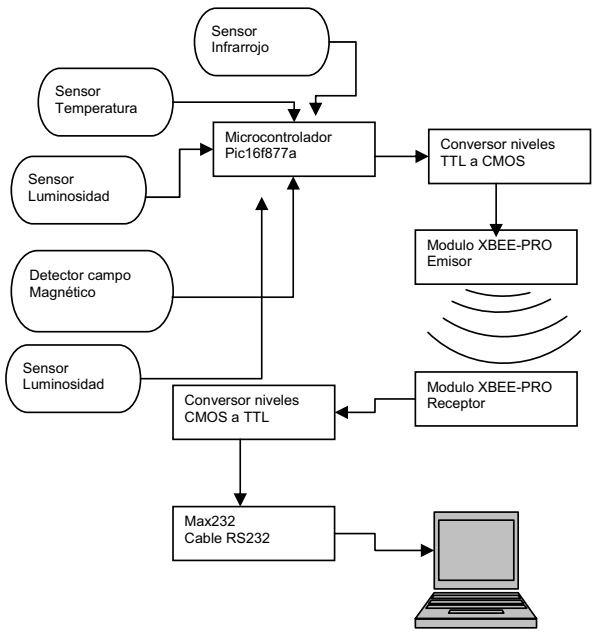


Fig. 8: Description of the basic functioning of the system

E. Tests

To test the communication's scope, there were placed two sensor units in different points of our University's campus that were separated 300 m from the receiver unit which was in the instrumentation lab, and separated approximately 100 m between them. The experiment was repeated multiple time in different locations and with different obstacles between sensor units and the receiver unit. There were also done tests in a highly congested urban zone, lowering the distances to 80 m for a stable communication. Finally, the results were the expected ones.

IV. Discussion and Conclusions

Based on the concept of virtual instruments and using the ZigBee comunication protocol, it has been developed a robust, stable and cheapest wireless network sensor. Moreover, this hardware – software platform have a huge capacity to be adapted on many diferents situations and ambients (houses, laboratoies, a woodland, a greenhouse ora n industry), allowing to develop monitoring and control systems for small and medium nacional industries in a quick, trustable, and cheapest way

The pincipal interest is to apply this network to monitoring physical variables on ecosystems and agricultura

This kind of network, developed in this work, allows a fast and cheap instalation, this is because the cables (wires) are very few, and get a huge elasticity to the network, in other words, it allow the basic units (sensored and receptor units) be installed easy in many places.

In the market, there are a wide range of sensors for diferents physical variables. These sensors can be plugged to de sensor units, and in this way, it make possible to build many types of virtual instruments with a high quality. This technology has a huge additional value, and for its cost, could by implemented in

the small and medium enterprises of the country, allowing them to increase their competitiveness in the globalized market.

One of the biggest advantages of software is the great versatility it has due to its implementation on Java language, which allows it to have a successful performance on diverse operating systems. Furthermore, being Java and all its used tools (for design and development) of free distribution, it's very possible the implementation of this kind of systems at a very low cost.

V. Recommendations and Development Fronts

It is very useful, added capacity to data storage (flash memory) to the unit sensor, self-identification of the sensors and reduces their energy consumption. This latter capability is very necessary for applications where media access is hindering

To optimize power consumption, is proposes a PIC microcontroller 16LF877A, which works with CMOS voltage levels and low consumption, making unnecesary the use of converters in CMOS TTL.

At the present, the Scientific and Industrial Instrumentation Group of National University of Colombia Sede Medellín, is working on the production of 20 sensor units, and in this way have a great system with many stations for tests. The reception unit is going to be plugged to a TINI Tiny InterNet Interfaces,[6], [7]) microcontroller, replacing the PC on each station. The major idea for this change is to test the efficient of this kind of sensor network.

VI. Acknowledgement

This article is result of the investigation Project "Design and Implementation of Virtual Instruments for Monitoring and Controlling Physical Variables Using Different Communication Protocols" which is financed by the Nacional Direction of investigation of National University of Colombia (DINAI, proyect #20101006586).

VII. References

[1] ZigBee Alliance. Wireless Control That Simply Works, [web] http://www.zigbee.org/ en/index.asp, [Último Acceso, 22 Agosto 2007].

[2] Nacional Instruments, *Our History* [web] http://www.ni.com/company/history.htm [Último Acceso, 22 septiembre 2007].

[3] House, R., *Choosing the right Sotware for Data Acquisition*, IEEE Spectrum, 1995 p. 24-34.

[4] Montoya, A., Aristizábal, D., Restrepo, ., Montoya, N., Giraldo, L., *Design and Implementation of Virtual Instruments for Monitoring and Controlling Physical Variables Using Different Comunication Protocol*, Aceptado para publicación en la revista de International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering avalado por la IEEE., 2006.

[5] Pulido, J., Serrano, C. , Chacón, R., *Monitoreo y Control de Procesos a través de la Web*. III Simposio de Control Automático. Habana, Cuba, 2001 p. 73 -78

[6] N. Montoya, L. Giraldo, D. Aristizábal, A. Montoya, *Remote Monitoring and Control System. of Physical Variables of a Greehouse through a 1-Wire Network,* International Conference on Telecommunications and Networking, University of Bridgeport, diciembre 10-20, 2005, Sobh, Tarek ,Elleithy, Khaled, *Circuits & Systems Advanced in Systems, Computing Sciences and Software Engineering* , **XIV**, 437 p., Springer, 2006, Hardcover ISBN: 1-4020-5262-6.

[7] A. F. Muñetón, J..M. Saldariaga A., A. Montoya, D. Aristizábal, *A Module for Monitoring and Remote Control of Physical Variables Using Mobile Devices,* International Conference on Telecommunications and Networking, University of Bridgeport, diciembre 10-20, 2005, Elleithy, K., Sobh, T., Mahmood, A., Iskander, M. Karim, M. (Eds.), *Circuits & Systems Advanced in Systems, Computing Sciences and Software Engineering*, **XV**, 489 p., Springer,2006, Hardcover ISBN: 1-4020-5260-X.

[8] Vidal, A., Pérez, F., Calle, A., Valdés, E., *Estrategias para una Instrumentación Virtual de Bajo Costo con Aplicaciones Médicas*., Memorias V Congreso de la Sociedad Cubana de Bioingeniería, Habana, Cuba, 2003., Artículo T_0060.

[9] Netbeans [web] http://www.netbeans.org/ [Último Acceso, 22 septiembre 2007].

[10] JFreeChart [Web] http://www.jfree.org/jfreechart/ [Último Acceso, 22 septiembre 2007].

Inter-Agent Communication Adaptations for Power Network Processes Simulation

Miroslav Prýmek and Aleš Horák
Faculty of Informatics
Masaryk University
Brno
Czech Republic

Abstract – **In the paper, we present the agent communication standards that are implemented in the Rice simulation system for management and control of processes conducted in an electrical power network. The system models the power network using the multi-agent approach, with the agent communication flowing through a 4-layer communication protocol.**

We summarize the main features of the Knowledge Query and Manipulation Language (KQML) which is used as the system's content language. These features are then confronted with specific requirements of the area of electrical power network simulation and specific adaptations directed to system optimizations are described.

I. Introduction

The quality of power supply is an essential requirement of all technical processes in today's society. Therefore the fault-tolerance of the power supply is very important and constantly evolving science branch which demands intensive research. Because of the geographic dissipation and big financial cost of the power systems facilities maintenance, the required reliability cannot be reached simply by redundancy. It is necessary to search for the new low-cost but effective means of the required reliability assurance.

In the current research, we have focused on the exploration of the possibilities of the usage of KQML as the knowledge distribution language in our power system simulator called Rice. The feature and architecture overview of the simulator prototype was published in [3] and [4] and is illustrated on Fig. 1.

The main purpose of the simulator is to model the costs of the maintenance of a power distribution network in comparison with the costs of the outage, failures and undelivered energy. [5]

Knowledge Query and Manipulation Language (KQML) has become de-facto standard agent communication language in the multi-agent systems. It is defined in specification [1] and specification revision proposal [2].

KQML represents a very versatile implementation-agnostic knowledge distribution language. Its versatility is based mainly on the separation of the two message layers – the KQML layer itself, representing the purpose and background of the message and the content layer representing the message data (coded in *content language* which can vary between messages and so use appropriate format for the particular knowledge).

The process of the simulator implementation has led to the discovery that the KQML features are versatile enough to build up an agent communication network suitable for the highly scalable, modular and distributed power system simulator. On the other hand, the universality of the KQML model limited the performance of the system and increased the complexity of the analysis.

To overcome these difficulties, we have proposed some knowledge distribution optimization which will be discussed in this paper.

II. KQML features

KQML features can be summarized in accordance to the *performative* groups. A performative is a well-defined keyword which declares the meaning and the purpose of the KQML message content (data, expression). KQML draft [1] defines these kinds of the performatives:

1. *basic informative performatives* (tell, deny, untell) – basic knowledge sharing
2. *database performatives* (insert, delete, delete-one, delete-all) – specific database operations
3. *basic responses* (error, sorry) – special responses for messages
4. *basic query performatives* (evaluate, reply, ask-if, ask-all, ask-about, ask-one, ask-all, sorry) – performatives for the agent knowledge base values retrieving
5. *multi-response query performatives* (stream-about, stream-all, eos) – performatives for special data handling
6. *basic effector performatives* (achieve, unachieve) – performatives dealing with goals and *actuators*
7. *generator performatives* (standby, ready, next, rest, discard, generator) – performatives for handling the series of responses
8. *capability-definition performatives* (advertise) – performative for declaring agent's capability to do something or communicate in some specific manner
9. *notification performatives* (subscribe, monitor) –

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 22–25.

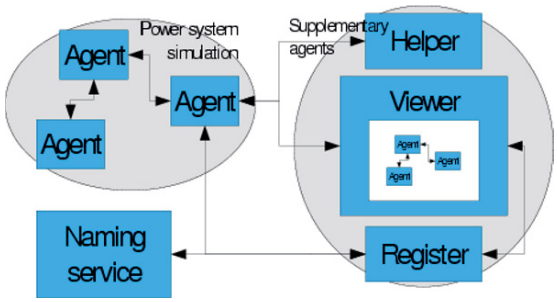


Fig. 1 - Power simulator architecture.

performatives for declaring that the agents wants to be constantly informed about changes of some other agent knowledge

10. *networking performatives* (register, unregister, forward, broadcast, pipe, break, transport-address) – performatives for networking and messages delivering
11. *facilitation performatives* (broker-one, broker-all, recommend-one, recommend-all, recruit-one, recruit-all) – performatives for the capabilities brokering and agents network building

As you can imagine from this list, the KQML specification is very versatile and defines all kinds of the agents communication which could be needed in a standard multi-agent application, even the distributed and dynamic (i.e. using agents hiring) one.

Complete implementation of a system which supports the whole KQML specification is a big task and there are only a few implementations (see e.g. [6]) of the KQML libraries. Complexity of such a communication system also means problematical performance and can constitute a significant bottle-neck in the high-performance systems. (For general discussion about performance of the Multi-Agent systems see [7]).

III. POWER SYTEM SIMULATION REQUIREMENTS

Rice system is developed as fully distributed and dynamical system consisting of the independent agents. The behavior of the system is defined strictly in the manner of defining the behavior of the agents of which it consists.

In our system we are simulating these processes:

1. *The energy flow and the transformation of its characteristics* – the topology of the distribution network vary by reorganizing load on the lines and switching the lines on and off. Amount of the energy distributed on every power line is limited by its maximal capacity. The lifetime of the particular power network components can be affected by their load.
2. *Cost of the maintenance of the power distribution network components* – each component holds an information about its condition and simulates the process of the maintenance – e.g. switching off during maintenance and building up the new distribution lane
3. *Components failures and their consequences*. Component failures can affect only the component condition or function or can have global consequences (outage).
4. *Outer events* - simulation of the natural disasters affecting power distribution network and user intervention in the system functions.
5. *Dynamic agents' behavior changes* – the user can change the agents' behavior *on demand,* even if it requires agent's program code change. Implementing this feature will also support interchanging of the *strategies* between agents.
6. *System function visualization* – graphical representation of the status of every system component and energy flow

The problem of the implementation of a multi-agent system which can meet all given requirements is definitely a problem of developing of the agent's knowledge interchange network which will support all functions needed for the simulation of the given processes. Which KQML features are needed?

IV. NEEDED KQML FEATURES

Energy characteristics are implemented as a set of the values in the agent's knowledge base. Being so, all *energy flow* can be implemented as a chain of the basic query-response messages (KQML performatives group 1 and 4 above).

Network topology is constituted by the query-response sender and receiver pairs. The agent which represents the source of the power is asked for the energy characteristics and the energy flow is established. Topology can be dynamically changed then by changing these pairs. To establish this lasting connection between agents we will need performative *subscribe* (group 9).

Because of the independent agent concept of the multi-agent systems, *maintenance cost* and *component failures* are private processes of the agents. These processes result in the status change of the agent and eventually information of all the subscribing agents about the change.

Outer events are implemented as an insertion of the particular knowledge into the agent's knowledge base. This requires the performative *tell.*

System function visualization is implemented in the stand-alone agent *Viewer*. This agent subscribes for the value changes of the knowledge which are to be visualized as the agent state (e.g. the actual power load of the line).

The last but not least function of the system is dynamically changeable agent behavior. This deserves *content language* in the KQML messages which is robust enough to transfer behavior code. The agent itself must be entirely or partly written in the programming language which supports dynamic code compilation (in our system we use Python language).

V. OPTIMIZATIONS

From the needed-feature list above, you can see that only a small portion of the KQML possibilities are to be used for the implementation of our power distribution system simulator. Realizing this in the process of the software analysis, we have proposed these optimizations:

1. To use multiple content languages appropriate for the purpose of the particular message, i.e.:
2. usage of the simple, easy-parseable content language for inserting and querying data of the agent knowledge base
3. usage of the higher programming languages (e.g. Python) for other purposes where the above-mentioned language will not be sufficient. These languages are general enough for any data structure definition (including behavior definitions – i.e. function definitions).
4. To develop a highly effective partial KQML parser with the best effectiveness in dealing with the most common messages, open for scalability and extendibility of the set of the KQML performatives understood.

Content languages

For interchanging complex data such as function definition we use Python programming language. This language makes possible ad hoc compilation and usage of the given code and gives us robust and general expression instrument. Usage of complex data is so rare that time-consuming code compilation process does not significantly affect general simulator performance.

However these higher-level languages are not suited for the exchange of the simple-structured data. With KQML we are not limited to use just one language for all purposes and rather we can use a language which is fittest for a given purpose. In our system, we divided message into two categories and use two expression languages for them.

The second language is used for setting and also querying agent's knowledge bases. We use simplest possible declarative language with expression in this form `label = value`, where label is a knowledge element label and value is a string representation of the given value. The special symbol '?' means that the value is not known or is to be asked.

The message content is constituted with one or more lines with the knowledge element/value pairs. This simple format can be parsed with regular expressions and all asked values ("?") can be automatically replaced with their respective values.

This declarative language in connection with the KQML concept of performatives is strong enough to express all operations on the knowledge base which are needed in our simulator:

`ask(voltage=?)`	which value is related with the knowledge labeled "voltage"? (which load is on the line)
`ask(power_free=10)`	is there (at least) 10 (kW) of a free distribution capacity?
`tell(voltage=10)`	load on the line is 10 (kV)
`untell(voltage=10).` ..or `tell(voltage=?)`	delete knowledge from the knowledge base
`tell(voltage=?)`	(as a reply to the *ask* message) the knowledge is not present in the agent knowledge base

Optimizing KQML parsing with a in-reply-to filtering

KQML messages semantic analysis is a difficult task. KQML messages can be nested constructing entities with a very complex meaning. To overcome this difficulty we used a concept of a partial KQML parser – there are defined canonical messages which are most frequent and such must be handled most effectively. To do this we use a concept of messages filtering according to the *in-reply-to* KQML property.

According to KQML specification, message should contain property *reply-with.* When answering to such a message, agent must send a message in which the property *in-reply-to* is set to a same value as the original message *reply-with* property.

By parsing the field *in-reply-to* first, we can identify the purpose of the message and we can suppose that the message is structured in accordance with the initial agreement (which should be part of the KQML *ontology* definition). Supposing this, the process of parsing and semantic analysis of the message can be simplified and significantly accelerated.

Using this concept we can construct KQML parser which handles most frequent messages much faster than a parser which will analyze the messages with no anticipatory knowledge. On the other hand, the universality of this parser is not affected – messages structured different than the ontology standard are analyzed with a full KQML analyzer.

VI. PERFORMANCE TESTS

To evaluate the proposed optimizations we have made some performance tests on our system. These tests measure the basic message passing functions performance.

The computer the tests run on was Apple Mini 1.6GHz with 1GB RAM. As the CORBA ORB was used omniORB version 4.0.2. All agents have run on this one computer and in the tests no LAN communication was involved.

Our prototype is written entirely in Python programming language, only Register and Viewer agents are written in Java.

All tests were run one hundred times. The arithmetic mean of the achieved performance values will be presented below.

CORBA performance: In this test the one cycle involved simple CORBA *one-way* function call with no arguments. With omniORB python library we managed to get 29 117 these cycles per second.

KQML structures passing: In this test the cycle involved KQML structure construction and passing it to the receiver. Performance was significantly lower – 15869 cycles per second.

Simple content language expression passing and decoding: In this test we evaluated the performance of the handling of the KQML message with the content in the simple declarative language described above. The test cycle involved construction of the KQML message with content `tell(testKnowledge=1)`, its passing to the receiving agent, decoding and inserting a value to the agent knowledge base. The average performance was 11274 cycles per second. As you can see, the performance has not been seriously affected in comparison with the previous test. This confirms that the application of the simple content language enables acceptable performance.

Knowledge exchange using simple content language: In this test we tried to simulate the process of the knowledge exchange which takes place in the real simulator system run. The test consisted of sending ten thousand messages in the form `ask(testKnowledge=?)` to the receiver and measuring time to the arrival of the last reply message. The average performance in this test was 3311 ask/reply message pairs passing per second.

The overall results of the tests are presented on figure 2. From the values gathered we can compute a gross estimation of the time-consummation of the respective message passing processes as shown on figure 3. Note that a CORBA call will be bypassed in a local-only system implementation and therefore the system performance will raise by one third in this scenario.

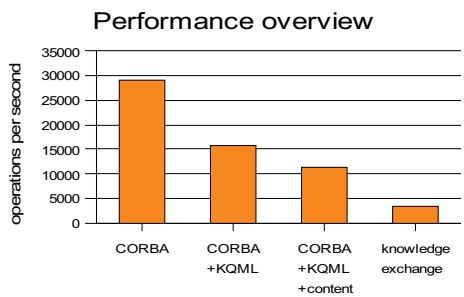


Fig 2: Performance tests results overview

VII. CONCLUSIONS

The proposed optimizations have been approved as reasoned and have led to a significant speed up of the inter-agent communication. Even if our prototype is written in a slow scripting language, the performance achieved is promising for the future implementation of the system in more effective programming language.

The main advantage of the proposed optimizations is that they do not break KQML specification conformance and can be aimed specifically at the bottle-neck of the inter-agent communication and also can be fine-tuned with respect to the frequencies of a particular KQML message types and structures.

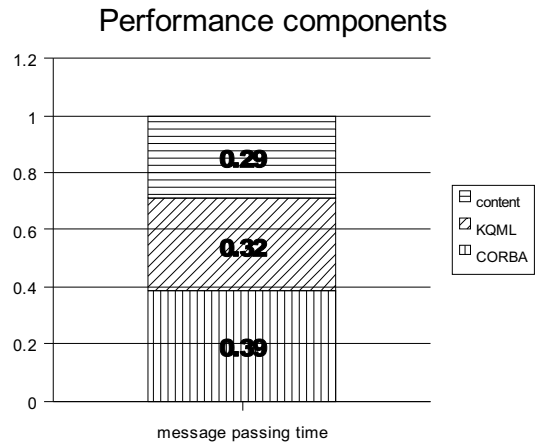


Fig 1: Components of the message passing time

REFERENCES

[1] The DARPA Knowledge Sharing Initiative, *Specification of the KQML Agent-Communication Language.* 1993. http://www.cs.umbc.edu/kqml/papers/kqmlspec.pdf

[2] Labrou,Y., Finin,T., *A Proposal for a new KQML Specification.* 1997. http://www.cs.umbc.edu/kqml/papers/kqml97.pdf

[3] Prýmek, M., Horák, A., "Multi-Agent Framework for Power Systems Simulation and Monitoring*."* In *Proceedings of ICICT 2005*. Cairo, Egypt : Information Technology Institute, Giza, Egypt, 2005. pp. 525-538. ISBN 0-7803-9270-1.

[4] Prýmek, M., Horák, A., "Design and Implementation of Multi-Agent System for Analysis of Electrical Power Networks." In *Proceedings of ELNET 2005.* Ostrava, Czech republic: VŠB TU, Ostrava, 2006. pp. 525-538. ISBN 80-248-0975-3

[5] Prýmek, M., Horák, A., "Modelling of Economical Aspects of a Power System Failure." In *Proceedings of PEA 2006.* Anaheim, USA : International Association of Science and Technology for Development, 2006. s. 61-66. ISBN 0-88986-616-3

[6] http://www.cs.umbc.edu/kqml/software/

[7] Lee, L., Nwana, H. S., Ndumu, D. T. & De Wilde, P. (1998), "The Stability, Scalability and Performance of Multi-Agent Systems," *BT Technology Journal* 16 (3), 94-103.

DC Motor Monitoring and Control System

Andrei Cozma

Applied Electronics Department, Technical University of Cluj Napoca, Romania
E-mail: Andrei.Cozma@ael.utcluj.ro

Abstract - **This paper describes the design and implementation of a system used for monitoring and controlling the operation of various types of DC motors. The system consists of two major components: a PC application and a hardware component controlled by a FPGA device. The role of the system implemented in the FPGA device is to acquire and process data related to the DC motor's operation, to control the operation of the motor and to implement a communication protocol for exchanging data with the PC application. The PC application provides to the user an interface for visualizing information related to the motor's operation and for interacting with the system. The PC application implements a speed controller, a speed and position controller and a series of automatic tuners for the speed controllers. The main advantage of the system is that it allows to automatically determine the control parameters for DC motors with different characteristics, and it allows to easily verify the performance of different controllers.**

Key words: FPGA, DC motor, Auto-tuner, PID controller.

I. Introduction

An indispensable component of any control system is the actuator. The actuator is the first system component to actually move, converting electric, pneumatic, or hydraulic energy into mechanical motion. The most common type of actuator is the electric motor. Electric motors are classified as either DC or AC, depending on the type of power they use. AC motors have some advantages over DC motors: they tend to be smaller, more reliable, and less expensive. However, they generally run at a fixed speed that is determined by the line frequency. DC motors have speed-control capability, which means that speed, torque, and even direction of rotation can be changed at any time to meet new conditions [1]. One technique for controlling a DC motor's speed is Pulse Width Modulation (*PWM*). In this system, power is supplied to the motor in the form of DC pulses of a fixed voltage. The width of the pulses is varied to control the motor speed. If the frequency of the pulses is high enough then the motor's inductance averages them, and it runs smoothly. One of the advantages that this system has is that the used power amplifier can be of the efficient class C type.

Together with the evolution and the improvement of Programmable Logic Devices (*PLD*) like CPLD (*Complex Programmable Logic Device*) and FPGA (*Field Programmable Gate Array*) a new class of control systems could be implemented for electric motors. The main advantage of the PLDs is that they offer to the digital circuit designer the possibility of changing functionality even after it has been built. A PLD can be programmed, erased, and reprogrammed many times, allowing easier prototyping and design modification [2].

In spite of all the advances in control over the past 50 years the PID (**P**roportional **I**ntegral **D**ifferential) controller is still the most common controller. Even if more sophisticated control laws are used, it is common practice to have an hierarchical structure with PID control at the lowest level [3]. The transfer function of the most basic PID controller is described by the following equation:

$$u(t) = K_p\left(e(t) + \frac{1}{T_i}\int_o^t e(\tau)d\tau + T_d\frac{de(t)}{dt}\right) \qquad (1)$$

where: K_P is the proportional factor, T_i is the integral time, T_d is the derivative time, $u(t)$ is the command value at the time t, $e(t)$ is the error value at the time t. PID controllers can be implemented in digital control systems. The transfer function of a digital PID controller can have more than one representation, depending on the chosen discretization method.

Because of the widespread use of PID control it is highly desirable to have efficient manual and automatic methods of tuning the controllers. Automatic tuning methods relay on a description of the process obtained from I/O data. In the great majority of cases, a process description is obtained by performing an experiment on the process. The most common types of experiments are the step response and the frequency response experiments. These experiments allow to determine a system's time domain characteristics (step response) and frequency domain characteristics (frequency response). Currently there are many methods available for determining the parameters of a PID controller from the time domain or the frequency domain characteristics, and each of these methods provides a controller designed to optimise a certain aspect of the system's behaviour, like the settling time, the overshoot, or the response time.

II. FPGA system design

The system implemented inside the FPGA device was designed taking into consideration the following functional requirements:

- Compute the DC motor's speed and position based on the signals received from the sensors placed on the motor
- Generate a PWM signal for controlling the DC motor's operation
- Control the speed of the DC motor using a PID control algorithm
- Exchange data with the PC application through a USB link according to a custom communication protocol. The system is capable to send information related to the motor's speed, position and command value, and to receive information about

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 26–31.

the speed controller's parameters, employed control type and motor command value

- Control the motor's operation according to the information received from the PC application

In Fig. 1 is presented the block diagram of the FPGA system with the connections and the data flow between the system's components.

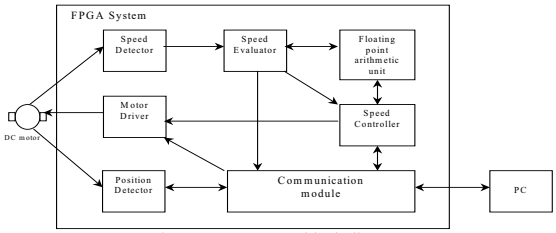


Fig. 1. FPGA system block diagram

A. Communication Module

The Communication module is responsible for the data exchange data with the PC application. It consists of three components: a USB communication component, a Receive Manager and a Send Manager. In Fig. 2 is presented a block diagram of the communication module.

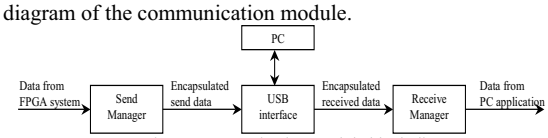


Fig. 2. Communication module block diagram

The USB communication component was developed starting from the Digilent Parallel Interface Module. The Digilent Port Interface is patterned after the EPP mode of the parallel port interface on an IBM PC type computer. The actual data transfer speed that can be achieved depends on the particular communications subsystem and firmware version being used [4].

The Receive Manager component decapsulates the data received from the USB module and dispatches it to the proper components. The data types that can be received are DC motor command value and speed controller parameters. The received data packets contain a 3 bits header for identifying the data type and 32 bits of data .

The Send Manager component acquires, stores and sends to the USB module the data received from the other modules. The sent data packets contain a 2 bits header for identifying the data type and 32 bits of data.

B. Speed Detector

The Speed Detector evaluates the speed of a DC motor from the signals generated by two Hall sensors situated next to the motor's shaft, at a 90° angle from each other. On the motor's shaft there is a disc with three magnets, which pass in front of the Hall sensors as the shaft is turning. The magnets are placed 120° apart from each other. The speed value provided by the speed detector represents the number of clock cycles generated by a 1MHz clock in the period of time necessary for the motor's shaft to perform a third of a complete rotation. This information is passed to the Speed Evaluator module that is responsible for computing the actual speed in rotations per second. The Speed Detector contains the following components:

- four Speed Counter modules which count the clock cycles between two successive passes of a magnet in front of the Hall sensors
- a filter which averages the data received from the four speed counters
- an integer to floating point converter which converts the filter's output to single precision floating point format

A Speed Counter module counts the number of clock cycles between two consecutive pulses generated by a Hall sensor. For each sensor two speed counters are used: one for counting the cycles between two consecutive rising edges and one for counting the cycles between two consecutive falling edges. Once all the four speed counters have generated a result, the results are taken by the speed filter which averages them and afterwards the average is converted to single precision floating point format. In Fig. 3 is presented the block diagram of the speed detector module.

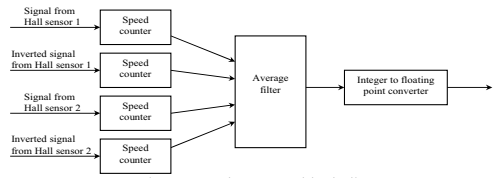


Fig. 3. Speed Detector block diagram

The signals generated by the Hall sensors do not have perfect edges, but oscillations occur at the beginning and end of the pulses. Because of this a debouncer circuit [2] was designed to receive as input the signal generated by a Hall sensor, and output an oscillation free signal. The heart of the debouncer circuit is a 2-bit comparator and a 50-bit serial shift register, with active-HIGH synchronous LOAD. The XNOR gate compares the shift register's serial input and output. When the shift register's input and output are different, the input data are serially shifted through the register. When the input and output of the shift register are the same, the binary value at the serial output is parallel-loaded back into all bits of the shift register. The basic schematic of the debouncer circuit is presented in Fig. 4.

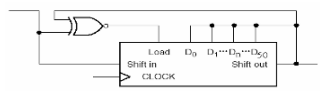


Fig. 4. Debouncer circuit

C. Speed Evaluator

The Speed Evaluator module computes, with the aid of the floating point arithmetic unit, the actual speed of the motor's shaft in rotations per second, based on the data received from the speed detector module. The speed is evaluated according to the following equation:

$$S = \frac{10^6}{3 * C}[rps] \quad (2)$$

where: S is the speed in rotations per second (rps), C represents the counts received from the speed detector. In Fig. 5 is presented a block diagram with the connections and the data flow between the Speed Evaluator module, the Speed Detector module and the floating point arithmetic unit.

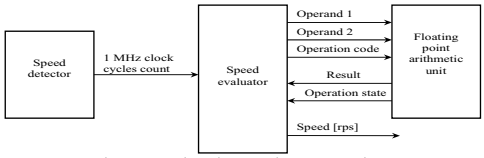


Fig. 5. Speed evaluator's interconnections

When a new speed value is computed it is sent to the PC application.

D. Position Detector

The Position Detector determines, based on the signals received from the two Hall sensors, the number of rotations and the direction in which the motor's shaft is turning. The accuracy of the position detector is a third of a full rotation. When a new position value is available it is sent to the PC application.

E. Speed Controller

The Speed Controller module implements a PID control algorithm, as described by the discrete transfer function:

$$u(z) = K_p\left(1 + \frac{Tz}{T_i(z-1)} + T_d\frac{z-1}{Tz}\right) \quad (3)$$

where: K_P , T_i and T_d have the same meaning as in equation (1) and T is the sampling period. The reference speed and the parameters of the PID controller are received from the PC application. When a new command value is available it is sent to the PC application. The new command value is also sent to the Motor Driver module to generate the appropriate PWM signal.

F. Floating Point Arithmetic

The arithmetic module was designed to perform arithmetic operations on single precision 32 bit floating point numbers. It implements the addition, subtraction, multiplication and division operations. In Fig. 6 is presented the block diagram of the arithmetic module.

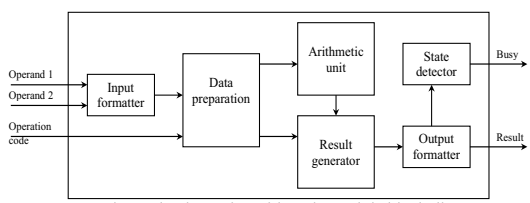


Fig.6. Floating point arithmetic module block diagram

According to the IEEE 754 standard [5] the format of the single-precision floating point numbers is:

$(-1)^s 2^e \times 1.f$ (normalized), when $e > 0$

$(-1)^s 2^{-126} \times 0.f$ (denormalized), when $e \leq 0$

In the cases when the exponent is positive, the "1" before the decimal point must be taken into account when performing arithmetic operations. The purpose of the Input Formatter block is to add this "1" as the most significant bit of the mantissa, so that it will be used in the following operations.

The Data Preparation module performs the following operations:

- formats the mantissas of the two operands according to the specifications of the arithmetic operation that must be performed, as described in [6]
- in case that an addition or a subtraction must be performed between two numbers with different signs the sign of the numbers and the operation code will be changed so that an equivalent arithmetic operation will be performed between two numbers with the same sign
- computes the exponent of the result
- evaluates the sign of the result

A data preparation operation takes one clock cycle to complete.

The Arithmetic Unit is responsible for performing the actual arithmetic operations. It takes as inputs the mantissas of the operands and outputs the mantissa of the result. The addition and subtraction operations take one clock cycle to complete, and the division and multiplication operations take 23 clock cycles to complete.

The Result Generator module was designed to take as input the sign, exponent and mantissa of a floating point number and, based on this data, to output a complete 32 bit floating point number. The input mantissa is received from the Arithmetic Unit and it is 24 bits long, unlike a standard mantissa which is 23 bits long. It must check the 24th bit of the mantissa, and if this bit is "1" to increment the exponent. The result generation operation takes one clock cycle to complete.

The Output Formatter module was designed to normalize a 32 bit floating point number according to the IEEE 754 standard. The number of clock cycles needed for the normalize operation to complete can vary between 1 clock cycle and 22 clock cycles. The fact that the normalize time is not constant is a disadvantage, but this solution was chosen in order to reduce the size of the component.

The State Detector module was designed to signal to the components that are using the arithmetic unit that an operation is currently in progress. The external components can monitor the signal output by this module to determine when the arithmetic module has finished the requested operation and the result is available on the output data bus.

G. Motor Driver

This module generates a PWM signal with a resolution of 10 bits and a frequency of 50 KHz. The fill factor of the PWM signal is received either from the PC application, or from the speed controller module.

III. FPGA SYSTEM IMPLEMENTATION

The system was implemented using the VHDL language and the *Xilinx WebPack 7.1i* CAD tool. The operation and the performance of the components were verified using the *ModelSim XE III 6.0a* simulation software.

In Fig. 7 is presented the system's top level schematic, while Table I provides a short description of the I/O signals.

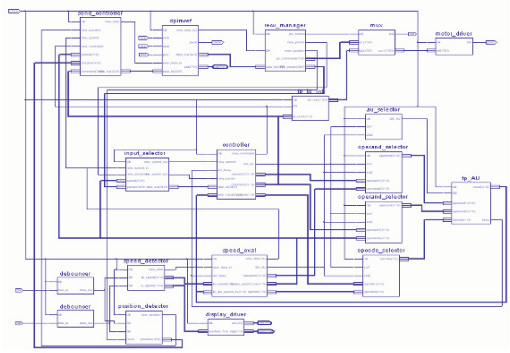


Fig. 7. FPGA system's Top level schematic

TABLE I
I/O SIGNALS DESCRIPTION

Signal	Mode	Description
Clock	Clock input	50MHz main clock input
Astb	Synchronous input	USB communication module Address Strobe.
Dstb	Synchronous input	USB communication module Data Strobe.
pwr	Synchronous input	USB communication module transfer direction control.
SA	Synchronous input	Signal from Hall sensor A.
SB	Synchronous input	Signal from Hall sensor B.
pwait	Output	USB communication module data synchronization signal.
pdb(7:0)	Bidirectional	USB communication module data in/out bus.
led	Output	Led D7 on development board. ON = motor control signal is generated by the PC application, OFF = motor control signal is generated by the FPGA PID controller.
pwm	Output	Pulse width modulated signal for motor control.
an(3:0)	Output	Anode control pins of the onboard seven segment display.
seg(7:0)	Output	Segment control pins of the onboard seven segment display.

Necessary hardware components: Digilent Nexys 400 board, Digilent Pmod HB5 board and power source, USB cable, DC motor with Hall sensors circuit, cable for connecting the DC motor to the HB5 board.

IV. PC APPLICATION DESIGN

The PC application was designed using an object oriented approach, and is made up of independent components which communicate with each other through messages, according to the observer design pattern [7]. This approach allows great flexibility because a component can be changed without affecting the others, and also the connections between the components can be changed very easily so that the system can be reconfigured dynamically. In Fig. 8 is presented the application's component diagram.

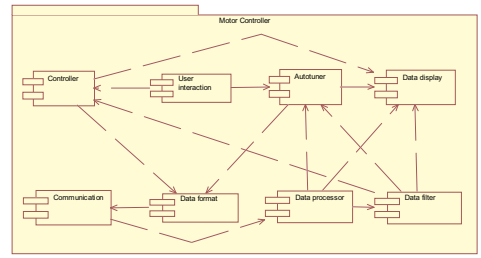


Fig. 8. PC application's component diagram

A. Communication

This component is responsible for managing the data transfer with the system implemented in the FPGA device according to the USB communication protocol's specifications. It uses the Digilent DPCUTIL Dynamic Link Library [4] to access the USB port.

B. Data Format

The Data Format component was designed to encapsulate the data that must be sent to the FPGA system according to a custom communication protocol. The objects that need to send data to the FPGA system must pass to this component a specific data structure which contains the type of the message that must be sent and the actual data. The following types of messages can be sent: command, reference speed, PID parameters and controller selection. The controller selection message is used for specifying if the speed controller implemented inside the FPGA will be used, or one of the controllers implemented inside the PC application will be used.

C. Data Processor

The Data Processor component decapsulates the data packets received from the FPGA system so that the data can be used by the other components.

D. Data Filter

This component is responsible for filtering the data related to the motor's speed received from the FPGA system, in order to eliminate the measurement noise. Currently it implements only a first order low pass filter with a filtering threshold, having the following transfer function:

$$y(t) = c \cdot y(t-1) + (1-c) \cdot u(t) \quad (4)$$

Where: $y(t)$ - filter output at time t, $y(t-1)$ - filter output at time $t-1$, $u(t)$ - filter input at time t, c - filter coefficient

The filtering threshold defines the maximum difference between the current filter input $u(t)$ and the previous filtered value $y(t-1)$ that must exist in order for the filtering to be applied. If this threshold is exceeded the filter outputs the

input value and the filtered value is set to that of the input value.

E. Controller

The Controller component contains a few types of controllers that can be used for regulating the operation of the DC motor.

The simplest controller is the manual controller, which receives as input the fill factor of the PWM signal and transmits this value to the FPGA system.

A speed controller is available for adjusting the motor's speed according to a specified reference speed. The speed controller uses a PID control algorithm.

A speed and position controller was designed to provide a controller capable of adjusting two parameters of the DC motor: average speed and number of rotations. The controller receives as parameters the number of rotations and the average speed with which the rotations must be executed. This is a cascade type controller, with a speed controller in the inner loop and a position controller in the outer loop. The purpose of the speed controller is to follow the speed trajectory generated by the position controller, using a PID control algorithm. The position controller generates a trapezoidal speed trajectory for the DC motor, based on the reference speed and the number of rotations that the motor must perform. It also supervises the error between the ideal position and actual position and corrects the speed trajectory using a PID algorithm, so that the real position gets as close as possible to the ideal position. The speed trajectory is computed so that it has a trapezoidal form, with equal rise and fall times, which expand for 15% (7.5% rise and 7.5% fall) of the total time needed to complete the specified number of rotations with the specified average speed. In Fig. 9 is presented the general shape of the speed trajectory, and the main parameters that characterize the trajectory.

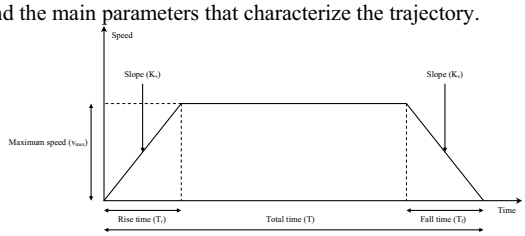


Fig. 9. Speed trajectory

The key parameter for computing the speed trajectory is the slope K_v, because it is used to determine the speed value at every moment of time during the rise and fall stages, and also the maximum speed. The slope is determined using the following equation:

$$v_{avg}=\frac{1}{T}\left[\int_0^{T_r}K_v\cdot t\,dt+\int_{T_r}^{T-T_f}K_v\cdot T_r\,dt+\int_{T-T_f}^{T}K_v\cdot(t-(T-T_f))\,dt\right]\Rightarrow K_v=\frac{v_{avg}}{c\cdot(1-c)\cdot T} \quad (5)$$

where v_{avg} represents the average speed and $c = T_r/T = T_f/T$.

F. Auto Tuner

The Auto Tuner component is used for automatically determining the parameters of the PID speed controller. The parameters are determined from the data obtained from a step response or a frequency response experiment. The Auto Tuner is responsible for generating the motor's command, according to the type of experiment to be performed. From the step and frequency response experimental data it is possible to determine the parameters of a PID controller using the following tuning methods:

- Ziegler-Nichols (PI, PID) – step response [8]
- Dahlin (PI, PID) – step response [8]
- Chein-Hrones-Reswick (PI, PID) – step response [9]
- Improved PI – step response [10]
- Åström and Häglund (PI, PID) – frequency response [10]

V. PC APPLICATION IMPLEMENTATION

The PC application was implemented using the C++ programming language and it runs under Windows operating system. It is based on a MFC (*Microsoft Foundation Classes*) single view framework. The implementation was done using the *Microsoft Visual Studio 6.0* development environment.

VI. PC APPLICATION PRESENTATION

In Fig. 10 is presented the application's main window. The main window contains two graphs which are used to display the evolution in time of the DC motor's speed and command values. In the topmost graph is displayed with red the motor's speed, in rotations per second, and with green the reference speed. In the bottom graph is displayed the command value. The command value represents the fill factor of the PWM signal used to control the motor and has values in the range 0:100[%]

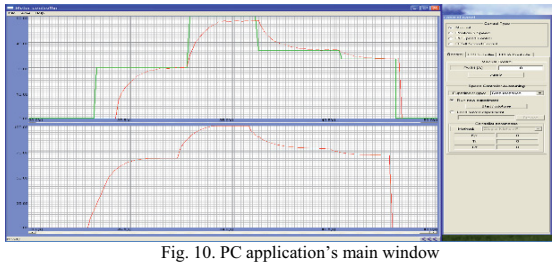


Fig. 10. PC application's main window

By right clicking inside a graph control a contextual menu opens allowing to change the graph's state. When a graph is running it constantly receives data from an external source and displays the received data. After the graph is stopped it retains all its content and allows horizontal scrolling, zooming and saving the displayed data to a file so that it can be reloaded at a later time.

From the *Control Panel* window it is possible to select the desired control type for the DC motor, to set the parameters of the different controllers and also to run auto tuning experiments. The following control types can be chosen:

- *Manual* – When this option is selected the *General* tab becomes active and it is possible to specify the DC motor's command value or to run speed controller tuning experiments.

- *Position & Speed* – When this option is selected the *PC Controller* tab becomes active and it is possible to specify the parameters for the speed and position controllers and to run position control experiments.

- *PC Speed Control* - When this option is selected the *PC Controller* tab becomes active and it is possible to specify the speed controller's parameters and to run speed control experiments.

- *FPGA Speed Control* – When this option is selected the *FPGA Controller* tab becomes active and it is possible to specify the parameters of the speed PID controller implemented in the FPGA device and to run speed control experiments using the FPGA controller.

The *General* tab from the *Control Panel* window allows the following operations to be performed:

- Change manually the motor's command

- Run auto-tuning experiments – There are two types of experiments available: step response and frequency response. It is possible to run a new experiment, or a saved experiment can be loaded. An experiment terminates automatically when enough data has been gathered, or the experiment can be stopped manually. After an experiment is complete, or after loading a saved experiment, it is possible to select one of the implemented auto tuning methods. Every time a new method is selected the PID controller's parameters are recomputed.

From the *PC Controller* tab it is possible to set the parameters and the reference values for the selected PC controller (*Speed controller* or *Position & Speed controller*), and also to start/stop the control process. During a position control cycle a dialog box is displayed, as shown in Fig. 11, to indicate the motor's average speed and number of rotations.

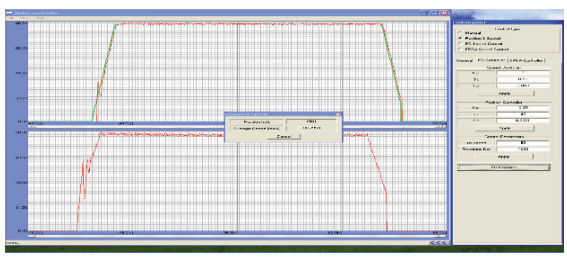


Fig. 11. Position control

From the *FPGA controller* tab it is possible to change the parameters of the speed controller implemented inside the FPGA system and also to set the motor's reference speed.

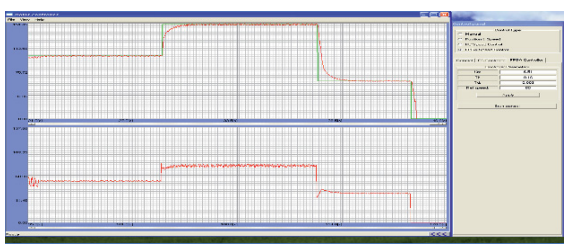


Fig. 12. FPGA control

VII. Conclusions

In this paper was presented the design and implementation of a system that is to be used for controlling and monitoring the operation of a DC motor. The main goal of this system is to provide to the user a series of alternatives for implementing the most suitable controller for a given application. The system offers some features which allow the automatic evaluation of the parameters for the used PID speed controllers. It must be mentioned that the performance of any tuning algorithm depends mostly on the quality of the used experimental data. In many cases the automatic tuning methods offer only a starting point in determining a controller's parameters, and the fine tuning of the controller must be done by the user, based on his knowledge and experience.

By comparing the performances of the speed controller implemented in the FPGA device and the speed controller implemented in the PC application, when using the same design parameters, it was noticed that the FPGA controller is working better than its PC counterpart. The reason for this performance difference comes from the fact that the response time is much better for the FPGA controller. The PC controller has to deal with the delays and the maximum sapling rate imposed by the operating system and the communication channel, while the FPGA controller responds without any delays to the changes in the motor's operation.

The main challenge in implementing the PC application was to develop an architecture that is able to provide minim response time and maximum sampling rate.

While implementing the FPGA system the biggest challenges were generated by the limited number of logic gates available inside the FPGA device, and by the signal propagation delays which become larger as the system increases in complexity and size. Also, as the size of the design increases, the development software tools require a considerable amount of time synthesise and implement the design.

References

[1] Thomson Delmar Learning, "Modern Control Technologies: Components and systems, Second edition"
[2] Thomson Delmar Learning, "Digital Design with CPLD Applications and VHDL"
[3] K.J. Astrom, T. Hagglund, "Revisiting the Ziegler–Nichols step response method for PID control", Journal of Process Control 14 (2004) 635–650
[4] "Digilent Port Communications Manual", www.digilentinc.com, 2005
[5] IEEE computer society: IEEE Standard 754 for Binary Floating-Point Arithmetic, 1985
[6] Jidan Al-Eryani, "Floating Point Unit"
[7] Erich Gamma, Richard Helm, Ralph Johnson, John Vlissides, "Design Patterns, Elements of reusable object oriented software", Addison Wesley, 1998
[8] A.Leva, C. Cox, A. Ruano, "Hands-on PID autotuning: a guide to better utilisation"
[9] Robert A. Paz, "The design of a PID controller", 2001
[10] Karl Johan Astrom, "Control System design", 2002

Web-Based Tele-Operated Control System of a Robotic Vehicle

Aneesh N. Chand
School of Electrical and Electronic Engineering
Fiji Institute of Technology
chand_an@fit.ac.fj

***Abstract*-This paper presents the design of a web-based tele-operated control system intended for office-wide geographically remote and real time control of a mobile robot utilizing networking technologies. A comprehensive account of the design paradigm for the web based control system of the robot is presented. Experiments are successfully carried out to verify the system implemented.**

I. Introduction

The explosion of the World Wide Web (WWW), the Internet, related web and internet technologies and their subsequent unprecedented growth has paved the way for the implementation of tele-operation of geographically isolated systems and has played a vital role in the advanced of tele-robotics.

The two most important reasons for this are that the Internet provides existent infrastructure used to connect two networks located in two different parts of the world. Secondly the standardization enforced by the IEEE and other standards organizations imply an ease of implementation and compatibility.

A lot has evolved from the time of the implementation of the world's first network, the ARPA by the Department of Defense (DoD) of the United States Military. Following that, the 1960s saw the introduction of mainframe computing and suggestions on packet switching technology.

The production of digital integrated circuits and the advent of personal computers and their widespread use in the following decades built the foundation for the networking phenomena that was soon to appear. The 1990s saw the introduction of technologies that would revolutionize the Internet for ever. The World Wide Web (WWW) was developed by Tim-Berners Lee at CERN (European Laboratory for Particle Physics) and the first graphics based web browser, Mosaic was conceived by Marc Andreeson and as a direct result of the birth of these technologies, there was faster-than-exponential growth in Internet usage in succeeding years.

In a short time thereafter, Internet and related web technologies rapidly found promising applications in robotics with the realization that they provided existent techniques and methods for geographically remote control.

Amongst the first significant works related to tele-operated robotics was the Mercury Project which consisted of an industrial robotic arm that was teleoperated via the WWW [1]-[3] and the development of the first Internet controlled mobile robot named *Xavier* [4] [5]. This robot is generally credited with being the world's first tele-operated mobile robot exploiting the full potential of internet-control. This robot was engineered by the Robotics Institute of Carnegie Mellon University and since it was commissioned online in 1995, it has received over thirty thousand (30,000) web requests and has traveled an equivalent distance of two hundred and ten kilometers (210km) as a result of remote tele-operation control. The robot used visual feedback to allow the remote operator real time control. Likewise, the manipulator of the Mercury Project was accessed by over 50,000 hosts via the Internet.

There is now a varied genre of robotic applications available today that utilize Internet technologies. Internet controlled lawn mowers, vacuuming robots and search and rescue robots have all been developed recently. A review of literature shows that robotic lawn-mover has been internet-controlled to allow a remotely stationed operator to mow lawn grass using a wireless local area network (WLAN) [6]. In addition, even a telerobotic robot dog [7] and a book browsing telerobotic manipulator have been contrived [8] while elsewhere, humanoids have also been controlled via networks [9].

The development of tele-operated systems following the pioneering works done expanded the functionality to even greater heights. No longer was tele-operation limited for research or for small scale experimental control applications. Further review of literature reveals numerous applications of internet control for nobale applications. Even an unmanned airship has been contrived by [10] that is controlled via the internet.

Within the context of e-robotics, the focus of attention and discussion in the cited works has been an abstracted overview of their accomplishments. This paper presents the generic design paradigm and implementation for the development of a tele-operated control system. The specific application chosen here for the demonstration of the design and implementation of the web control system is the real time control for a mobile robot.

The paper is organized as follows. Section II provides an overview of the operational framework overview of the system. The design of the web based control system is discussed in Section III. Experiments and concluding remarks and recommendations for further work are given in Sections IV and V, respectively.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 32–36.

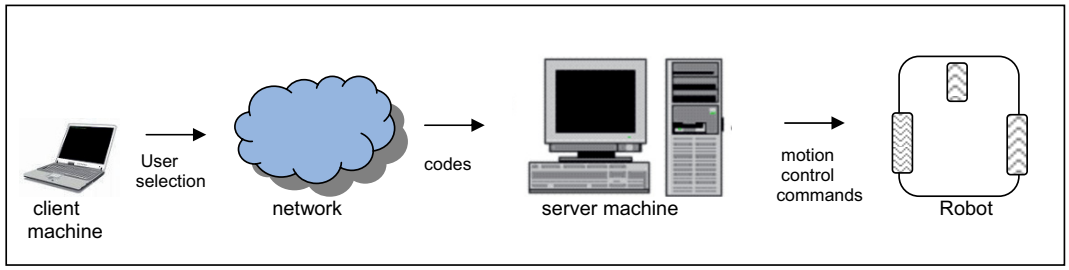


Fig. 1. Framework of the web based control of a robot

II. Design Overview

The exhibit of Fig. 1 illustrates the operational framework of the web based control system of a computer controlled robotic vehicle. A remotely stationed operator uses a client machine with Java enabled web browser to connect to the server machine using the TCP/IP and HTTP protocols. The server is directly connected to a robot. Selection of a motion control command from the web pages and the subsequent submission to the server triggers the selected robot motion. The client and server machines are connected using Ethernet technology. The following sections provide the design methodology and implementation of the system.

III. Design and Implementation

A. Web Pages

Web pages form the human-machine interface (HMI) between the robot and a remote operator. A combination of Hyper Text Markup Language (HTML) and Java applets are used in the development of the web pages. HTML is required, by default, in the construction of any web page while Java applets are used for other important reasons.

The single most important reason for using Java is that Java supports cross-platform deployment compatibility and this is the precise reason for the development of the Java language itself. When the Green Team of Sun Microsystems set out to write a programming language, the primary intent was to attain cross platform compatibility. This was achieved by changing the compilation output of Java. Conventional programming languages such as C/C++ compile into machine language which is then executed on a machine. Since machine language is machine-specific, this restricts the deployment of the C/C++ executable on the developer machine itself. Java circumnavigates around this dilemma by compiling not into machine language but into Java byte code. The Java byte code then runs on the Java Virtual Machine (JVM) and there is virtually a JVM available for download for all types of machines from Sun Microsystems. This process is illustrated in Fig. 2. This technique then guarantees that a Java application may be executed on any platform or machine irrespective of the platform or machine on which it was

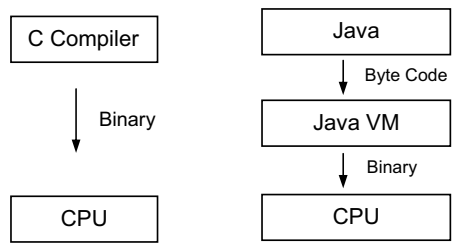


Fig. 2. Use of JVM as the compilation output to attain cross platform compatibility in contrast to compilation output of traditional C compilers

developed provided the running machine has the appropriate JVM installed. Since the operating system of client machines cannot be preempted, Java is therefore used to create the Java applets for cross-platform compatibility. A secondary reason for the deployment of Java applets is that the Java applet utilization augments security. For the reasons that server-side scripts are executed in the server, this implies the hiding of the filename, location and source codes of the scripts from the HTML codes of the web pages which prevent malicious users from tweaking and tampering.

A java enables web browser (Internet Explorer, Netscape Navigator or Mozilla FireFox) is then used to display the Hyper Text Markup Language (HTML) language and the embedded Java applet as a web page. The web browser on the client machine contacts the web server over the network, requests for information, receives the requested information and displays the results on the client screen. This is the only component of the tele-operation control design that is application specific as the logic of the application dictates the general design of the web pages. The only requirement of the web browser is that it must be Java-enabled to allow the deployment of the Java applets in the web pages.

B. Server Configuration

Linux (RedHat distribution version 9.0) was used as the operating system for the server machine for several reasons including greater control of the software. In addition, the server installation of Red Hat Linux provides all the

TABLE 1
MOVEMENT CODES AND CORESSPONDING MOTOE ACTIONS

Package	Description	Configuration File
`httpd`	Web server	`/etc/httpd.conf`
`xinetd`	FTP server	`/etc/xinetd.conf`
`bind`	DNS server	`/etc/named.conf`

necessary packages needed in the configuration and hosting of a server. These include the Apache web server, the File Transfer Protocol (FTP), the Dynamic Host Configuration Protocol (DHCP) and the Domain Name Server (DNS) server packages. Table 1 lists the relevant server packages that are distributed together with RedHat together with their respective configuration files that need to be appropriately configured to host the server.

Apachee was chosen as the web server as it is the world's most frequently used web server, is open source and the availability and quality of technical documentation and support relating to its use far surpasses those of propriety web server programs.

The server was hosted on a Class C private address and sub netted to allow the required number of hosts.

C. Server Side Dynamic Content Generation

A number of server-side technologies provide the methodology of dynamic web content generation. These are called server-side scripts as they reside in the server and are executed on invocation by a client browser. We first provide an overview of some of the available server side scripting languages as this will determine the selection of the appropriate language. Some common server side scripting languages are:

- Common Gateway Interface
- Hypertext Processor
- Java Server Pages
- Active Server Pages

Common Gateway Interface (CGI) is the traditional method of interaction between a server and back-end programs for the generation of dynamic content. A CGI script may be written in numerous scripting languages. These include, but are not limited to C/C++, Perl and Python.

A contribution of the open source community, Hypertext Preprocessor (PHP) is the open source extension of CGI. However, this requires explicit installation of PHP.

Java Server Pages (JSP) is a technique developed by Sun Micro Systems to generate dynamic HTML pages using the Java programming language. Use of JSP however sustains overhead as the Java language is often interpreted. Hence this technique is not favored in a real time control system where instantaneous reaction of the tele-operated hardware is required for minimal latency.

Active Server Pages (ASP) is Microsoft Corporation's contribution, The Visual Basic language is used for the dynamic content generation phase.

All the web content generation technologies stated above provide similar functionality. The selection of one dynamic web content generation methodology from the above techniques is therefore based mainly on personal preference (open source or propriety) or the merits of one technology over the others. The language selected for the work described here was CGI for reasons given in subsequent sections.

CGI is a communications specification designed to allow communication between external programs and a web server. More specifically, CGI is used to supply input from a remote client to a web server over a network and return dynamic content generated by the server back to the client machine to be displayed within the web pages. CGI scripts may be written in several programming languages. Recent years have seen the development of new technologies such as JavaScript, ActiveX and PHP which may be substituted in place of CGI but CGI is still by far the most widely used technology. Moreover, the CGI standard is the most compatible with current day browsers requiring no plug-ins to be installed, as is required by other standards such as Java and PHP.

A C++ CGI script was therefore written to accept input from the client machine. The selection of a motion command by a user and the invocation of the submission of the web page containing the user selection supplies the input to the CGI script. Each motion option command is represented as a numerical code in the Java applet and it is the code that is the actual input to the CGI script. There are in fact two methods of input submission to a CGI script, the GET and the POST method. The difference in the two methods is the visibility of the input in the URL after invocation of submission. Given security was paramount; the POST method of input submission was implemented. Upon receipt of a motion control code, the script writes the code to a text file.

D. Control Software

The control software was also written in the C++ language utilizing the GNU g++ compiler that is distributed together with Red Hat. Upon commencement of execution, the control software reads in the text file containing the motion control command code, applies elementary decision making structures to work out what application of the input signals to the motor controller circuits would achieve the user selected robot motion and correspondingly generates the logic signals through the parallel port. The motor controller circuits of the robot are interfaced to the parallel port (lpd) of the server.

The control software is first configured as a system daemon in Linux. A daemon is simply an executable in Linux that is de-attached from the console terminal to allow automatic execution. The control software is then configured to automatically commence execution whenever the server boots up. This ensures continual operation of the control software.

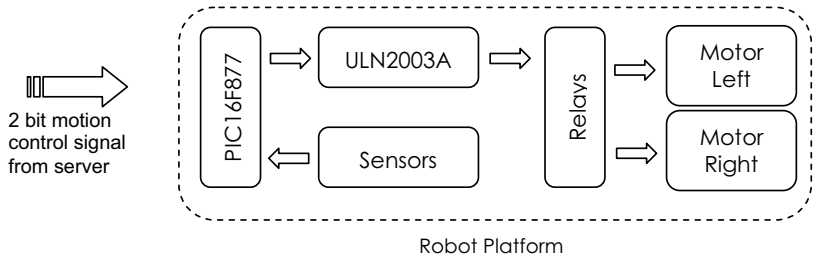


Fig. 3. Operational framework of robot electronics.

E. Security

Security is an important issue to consider in the implementation of any network. The methods of security chosen here was the incorporation of user authentication. User authentication is the first step of the security process in which the remote operator connects to the server machine using the client browser and supplies correct credentials which are verified. An affirmative match results in a successful login. In the current form, the subject robotic vehicle is hosted on a private and independent network. Hence user authentication was the only security feature utilized. Taking into account any future development in which the robot may be placed on the open network, firewalls and access control lists (ACLs) will be used for improvement of security.

F. The Robot

A simple robotic vehicle based on the design of [11] was developed for proof of concept of the web based control system. A differential drive system was constructed with the two driving wheels mated with high-torque, low speed permanent magnet direct current (PMDC) motors. The control circuitry of the robot encompasses a dc motor controller circuit, microcontroller board, infra red sensor for obstacle for obstacle detection and a simple voltage regulator circuit. The motor controller circuit is constructed using a monolithic, high voltage, high current Darlington pair current driver (ULN2003A) and four (4) Single Pole Double Throw (SPDT) relays which are used to control the rotational direction of the dc motors by controlling the direction of the flow of current through the motors. The controller of the robot is a PICmicro® microcontroller (PIC16F877) eight bit CMOS microcontroller with built in EPROM. The microcontroller board is adapted on the circuit schematics provided by Microchip Technology [12]. This microcontroller receives the motion control commands from the parallel port of the server machine, generates and transmits logic signals through the ULN2003A to the motor controller circuit for the attainment of the robot locomotion.

In the current form, the robot is hardwired to the parallel port of the server machine via a unidirectional parallel port interfacing card. Depending on the motion control command submitted by a remote operator from the client machine, the control software generates corresponding 2-bit codes through the parallel port of the server machine. The microcontroller on board the robot receives these codes and controls the movement of the robot accordingly. Fig. 3 illustrates the nomenclature of the robot electronics.

IV. Experiments

The validity of the web based control system for the mobile robot has been verified by implementing the system as described above in a 100Mbps LAN.

Amongst others, the major focus of attention in the experimentation was the latency or time lapse between the instant a user issues a motion selection command and the triggering of the respective robot motion. From the experiments, it was noted that are negligible communication delays and packet losses in an office-wide LAN. Accordingly, the mobile robot has been controlled remotely numerous times for basic tasks and demonstrations. The trials were repeated a significant number of times and on each occasion the robot performed as expected.

As a result of these experiments, it is verified that that the web based control system is sufficient enough to teleoperate a mobile robot in real time without sustaining substantial communication delays and packet losses using a moderately sized LAN. As such, possible applications include tele operation of indoor robots, such as inter office mail delivery or security robots, in building or office-wide LANs of limited physical size and geographical scope.

V. Conclusion

This paper has presented a comprehensive account of the design concept of the web based control system of a mobile robotic vehicle.

Different tiers in the system have different functionalities with web pages encompassing java applets forming the front end interface with which user select motion control commands, a sever accepts the motion control commands and

then transmits corresponding motion codes to the robot which finally performs some useful task.

Intended application include of the web based control system is the controlling of inter office robots for basic tasks such as mail delivery or collection in an office scenario where there would be a number of geographically distributed end users of the robot.

Current inhibitions include the limited operating range of the robot attributed to the hardwired system. Further improvements to the system will involve placing the robot on the open Internet, providing visual feedback to the operator by means of a wireless network camera and incorporation of a wireless connectivity between the server machine and the robot.

References

[1] Goldberg, K., Mascha, K., Genter, M., Tothenberg, N., Sutter, C., and Wiegley, J. (1995) Desktop Operation via the World Wide Web. *Proceedings of the IEEE International Conference on Robotics and Automation, pp 654-659.*

[2] Goldberg, K., Mascha, K., Genter, M., Tothenberg, N., Sutter, C., and Wiegley, J. (1996) A Telerobotic Garden on the World Wide Web. *SPIE Robotics and Machine Perception Newsletter, 5(1).*

[3] Goldberg, K., Mascha, K., Genter, M., Tothenberg, N., Sutter, C., and Wiegley, J. (2000) The Mercury Project: A Feasibility Study for Internet Robots. *Robotics and Automation Magazine,* IEEE Volume 7, Issue 1, pp. 35-40.

[4] Simmons, R. (1998) Xavier: An Autonomous Mobile Robot on the Web. *IEEE International Conference on Intelligent Robots and Systems (IORS):Workshop on Web Robots.* Victoria, Canada.

[5] Simmons, R., Fernandez, J. L., Goodwin, R., Koenig, S. and O'Sullivan, J. (2000) Lessons learned from Xavier. *Robotics and Automation Magazine,* IEEE Volume 7, Issue 2, pp. 28-32.

[6] Potgieter, J., Bright, G., Diegel, O. and Tlale, S. (2002) Internet Control of a Domestic Robot using a Wireless LAN. *Proceedings of the 2002 Australasian Conference on Robotics and Automation. Auckland, New Zealand, pp. 27-29.*

[7] Ferworn, A., Rouque, R. and Vecchia, I. (1999) MAX: Wireless Teleoperation via the World Wide Web. *Proceedings of the Canadian Conference on Electrical and Computer Engineering, Alberta, Canada.*

[8] Tomizawa, T., Ohya, A. and Yuta, S. (2002) Book Browsing System using an Autonomous Mobile Robot Teleoperated via the Internet. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp 1284-1289,* Switzerland.

[9] Sian, N. E., Yokoi, K., Kajita, S., Kanehiro, F. and Tanie, K**.** (2002) Whole Body Teleoperation of a Humanoid Robot. Development of a Simple Master Device using Joysticks. *Proceedings of the 2002 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems, pp 2569-2574,* Swizterland.

[10] Ramos, J., Maeta, S., Mirisola, L., Bueno, S., Bergerman, M., Faria, B., Pinto, G. and Bruciapaglia, A. (2003) Internet–based Solutions in the Development and operation of an unmanned robotic airship. *Proceedings of the IEEE, Volume 91, No. 3,* pp 463-473.

[11] Chand, A. N. and Onwubolu, G. C. (2006) Development of a Joystick based Control for a Differential Drive Robot. *Innovative Algorithms and Techniques in Automation, Industrial Electronics and Telecommunications*, Springer Netherlands, pp37-41.

[12] AN732 Implementing a Bootloader for the PIC16F87X, Microchip Technologies Inc, 2000, USA.

Number Plate Recognition Using Analytical Fourier Mellin Transform

Anshul Mittal
Hughes Systique

Mayank Sharma
Hughes Systique

Abstract

An automated, robust and fast vehicle license plate recognition system has become a essential necessity for traffic control and law enforcement in the present scenario. With increasing volumes of traffic being added to Indian roads every year and rough weather conditions prevalent, the challenge has become tougher. Here, we present a novel OCR-based method for number plate recognition using the Analytical Fourier Mellin transform which is fast and yields better results as compared to basic Fourier Mellin Transform.

The overall system uses an intelligent but cost-effective camera, which also uses minimal space, because the processor and the sensor are integrated in one package. They are more robust and reliable in outdoor environments, they require less maintenance and they are well suited for cost sensitive applications. Traditional correlation methods have to calculate the correlation over all possible rotations to determine the rotation angle between the template and query number plate character image when this angle is unknown. In this paper we achieve the alignment by Analytical Fourier-Mellin transform (AFMT), which can quickly align two images without re sampling errors and the high computation cost involved in conversion from log-polar to Cartesian. The Signature Technique has been used for the Number Plate Localization and Peaks to Valleys method have been used for character segmentation. As in the Indian Scenarios, the character distortion caused by dirt particles, image blurring or shadows that partially cover the license plate may cause problems. Thus, the segmentation is preferably parameterized to be sensitive to spaces. To further speed up the system, the property of symmetry of the vehicle rear and front parts are uniquely used to make the process of localization/identification of the license plate quicker.. To ensure the identified character is not a caption a syntax analysis is subsequently performed. The experimental results with number plate correlation databases show that the proposed algorithm is effective.

Index Terms – number plate matching, Analytical Fourier-Mellin transform, Signature technique, Peaks to Valleys

1. Introduction

The number plate recognition has been a very effective and useful technique that has been deployed in the recent past for automated parking lots or for monitoring and tracking the vehicle flow. Robustness, speed and accuracy is still a challenging problem especially in rough conditions where dirty, captioned or non standard number plates are still existant.

The scheme that is frequently incorporated for number plate recognition is correlation-based number plate matching. Here, the template and query number plate images are spatially correlated to estimate the degree of similarity between them. Correlation-based techniques are attractive due to features such as shift invariance and ability to accommodate in-class image variability. One problem is when the rotation of the query with respect to the template are not known, the correlation must computed over all possible rotations, which is computationally very expensive.

Even when using the often used standard Fourier Mellin transform [1]-[3], since it being based on the discrete log-polar coordinates, the numerical conversion from Cartesian to log-polar coordinates of the discrete image is required, and will bring significant re sampling error, which severely interferes with the resulting transformation parameters. Furthermore, the error introduced by the interpolation may reduce the accuracy of registration, whereas the process of high-order re sampling may produce higher computation cost.

In this paper we have proposed AFMT technique aiming at finite discrete images that can eliminate the conversion between the two coordinate spaces, avoid the disadvantages suggested above. As for the image capturing the camera may itself include the processor for managing image data, and executing the license plate recognition software. Accounting for noise, low contrast and blurring of the image, and performing requisite pre-processing of the image obtain a higher quality baseline image for registration. A proposed localization procedure then would include having a means to provide an initial projection profile, computation of which would be limited by use of a proposed symmetry based localization procedure, and a filtering threshold across vertical lines of pixels in the baseline image. Then, segmentation of characters depicted in the baseline image is done based on projection along horizontal axis of the baseline image to identify positions of the characters. The memory includes database for predetermined license plate codes and the standard syntax for accurate license plate detection.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 37–42.

The next section presents our proposed scheme used from the point of image processing to the point of localization of the license plate. Subsequent four sections deal with the proposed character recognition scheme. Section 2 of this paper presents the model for the misalignment between images based on their geometric transformation. Section 3 and 4 review the standard and analytical Fourier-Mellin transform underlying our algorithm, respectively. Section 5 and 6 present our registration algorithm, including approximation, implementation, and practical considerations. Section 7 presents experimental results of the proposed algorithm based on different criteria. Sections 8 and 9 make some concluding remarks about our applied algorithm

2. Initial Image processing

2.1 Initial image processing

The input image is initially processed to improve its quality by normalizing the image with some threshold which helps to offset the differences resulting form changing environmental and lighting conditions, mechanical damages to the license plates. It increases the contrast for obtaining a suitable binary (black and white) image from the grey image obtained by camera for license plate identification. It involves

(1) Conversion of gray intensity image into a binary image by application of Canny method of segmentation [18]. It finds edges by looking for local maxima of the gradient of the B & W Image.
(2) Dilation using the linear morphological structuring element orthogonal to each other: Using this structuring element, holes and spots are filled, to make license plate and character localization easier.
(3) Sub-sampling: Averaging filter is applied, based on a fixed standard/threshold, in order to reduce the number of pixels herein in the image, leading to lesser computational complexity.

2.2 License plate localization using signature technique and symmetry based localization

Signature Technique. The "signature" is a characteristic sequence of minimum and maximum in the brightness function calculated on a single row of input image that crosses a license plate. This sequence is characterized by some regular intervals. To localize the license plate, rows of the input image selected by centroid criterion are checked. If there is a part of a row that looks like a "signature" this part is marked. Based on this, more than one license plate candidates may be obtained.

Symmetry based localization. The object boundaries of different distinct objects (same pixel value connected components) are evaluated to assess the candidate having the highest magnitude area. This should be the car boundary. Now, owing to the inherent symmetry in the front and rear parts of a vehicle where the centroids of license plate and the car front/rear lie close to each other on the horizontal axis, the centroid of the object image is evaluated wherein computation is only required from one boundary of car to the point identified by coordinates of centroid + half license plate length + some compensated length on the horizontal axis. This reduces computational costs considerably.

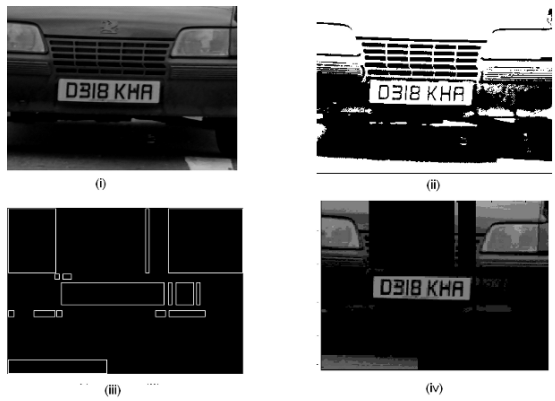


Fig.1 Stages of image localization

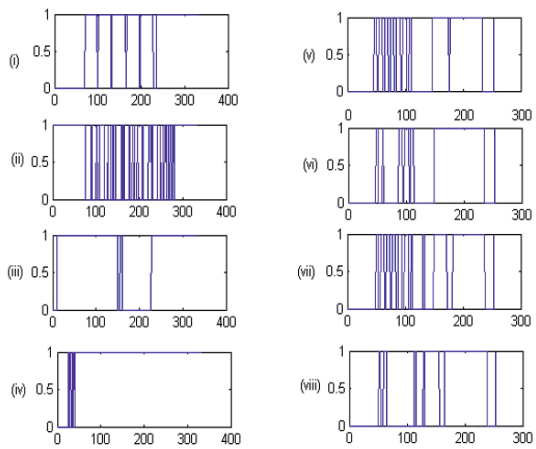


Fig. 2 Signature technique for fine localization of license plate. Fig.2 (i)-(iv) denote peaks for increasing y coordinates of image, while (v)-(viii) denote peaks for increasing x coordinates of image (taking top-left corner of image as origin). Note the high frequency of peaks and valleys for (ii) and (vi). These lie in the region of the license plate location

Character segmentation by "Peaks and Valleys" method. After the previous fine localization process, the license plate region should correspond to the maxima obtained for points of "valleys" surrounded with "peaks"

in all computed projections, and vice versa. Character segmentation is based on "peak-to-valley" function which is calculated on vertical edges of the candidate image (vertical edges dominate in license plates). Horizontal edges may also be employed for better plate localization, and segmentation of license plate region of interest be performed. These regions are selected for further processing. The character distortion due to dirt particles, blurring of image or shadows may cause problems

While sophisticated segmentation algorithms such as Hidden Markov Models, which are fully used for handwriting recognition or speech recognition, may perform better, here a simple segmentation based on projections to the baseline is advantageously employed to minimize execution time.

2.3. Character recognition

License plate recognition program device provides segmentation of characters by employing a projection along a horizontal axis of the base-line image to identify positions of the characters. The segmented characters are correlated with the stored database of templates using the below described Analytical Fourier Mellin Transform combined with phase-only correlation for image processing/editing.

3. General Image Transformation

Consider two functions denoted by *f* and *g*, representing a gray-level image defined over a compact set of R^2, whose misalignment can be modeled by a four-parameter geometric transformation that maps each point (x_f, y_f) in F to a corresponding point (x_g, y_g) in F_1 according to the matrix equation

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha\cos\beta & \alpha\sin\beta & -\Delta x \\ -\alpha\sin\beta & \alpha\cos\beta & -\Delta y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} \quad (1)$$

Equivalently, for any pixel (x, y) it is true that

$g(x, y) = f(\alpha(x\cos\beta + y\sin\beta) - \Delta x, \alpha(-x\sin\beta + y\cos\beta) - \Delta y)$, (2)

Where Δx and Δy are translations, α is the uniform scale factor, and β is the rotation angle. According to the translation, reciprocal scaling and rotation properties, the magnitudes of the Fourier transform of these images are invariant to translation but retain the effect of rotation and scaling, as follows

$$|G(u,v)| = \left| F\left(\frac{u\cos\beta + v\sin\beta}{\alpha}, \frac{-u\sin\beta + v\cos\beta}{\alpha} \right) \right| \quad (3)$$

Where $F(u,v)$ and $G(u,v)$ are the Fourier transforms of $f(x, y)$ and $g(x, y)$ respectively.

4. Standard Mellin Transform

The standard Fourier-Mellin transform of *f* in form of the polar coordinate system is given by:

$$M_f(k,v) = \frac{1}{2\pi} \int_0^{\infty} \int_0^{2\pi} f(r,\theta) r^{-iv} e^{-ik\theta} d\theta \frac{dr}{r} \quad (4)$$

For $\forall(k,v) \in Z \times R$. *f* is assumed to be summable over $R^* \times S^+$ (R^* denotes the multiplicative group of positive and non-zero real numbers, and S^+ the unit circle of the plane R^2) under the measure $d\theta \frac{dr}{r}$ _ , i.e.

$$\int_0^{\infty} \int_0^{2\pi} \left| f(r,\theta) r^{-iv} e^{-ik\theta} \right| d\theta \frac{dr}{r} = \int_0^{\infty} \int_0^{2\pi} \frac{1}{r} f(r,\theta) d\theta dr < \infty \quad (5)$$

Since *f* is positive. Note that the modulus of the Fourier-Mellin transform is rotation and scale invariant.

5. The analytical Fourier-Mellin transform

The FMT is defined for functions satisfying Eq. (5). However, in general, gray-scale images cannot meet this requirement and be assimilated to such functions since, in the vicinity of the center of co-ordinates, *f* is equivalent to the gray value of the centroid, which is generally different from 0. As a result, *f* is not summable and the integrals in (4) diverge. Ghorbel [6] proposed to compute the FMT of $f\sigma(r,\theta) = r^{\sigma} f(r,\theta)$ with $\sigma > 0$. The FMT of $f\sigma$ is called the Analytical Fourier-Mellin Transform (AFMT) of *f* and can be written in this way:

$$M_f(k,v) = \frac{1}{2\pi} \int_0^{\infty} \int_0^{2\pi} f(r,\theta) r^{\sigma - iv} e^{-ik\theta} d\theta \frac{dr}{r} \quad (6)$$

with $\sigma > 0$ for $\forall(k,v) \in Z \times R$. M_{f_σ} is assumed to *be summable* over $Z \times R$.

Let *g* denote the rotation and size change of a gray-level image *f* through angle $\beta \in S^+$ and scale factor $\alpha \in R^*$, i.e. $g(r,\theta) = f(\alpha r, \theta + \beta)$. The AFMT of g is

$$M_f(k,v) = \frac{1}{2\pi} \int_0^{\infty} \int_0^{2\pi} f(\alpha r, \theta + \beta) r^{\sigma - iv} e^{-ik\theta} d\theta \frac{dr}{r} \quad (7)$$

For $\forall(k,v) \in Z \times R$ and performing a simple change of the variables *r* and θ gives the following relations:

$$M_{g_\sigma}(k,v) = \alpha^{-\sigma+iv} e^{-ik\beta}\ M_{f_\sigma}(k, v) \quad (8)$$

For $\forall(k,v) \in Z \times R$. The relation in Eq. (8) makes the AFMT appropriate for extracting features that are invariant to scale and rotation changes [7], [11]-[15]. However, the usual modulus-based FMT descriptors are no longer invariant to scale because of the $\alpha^{-\sigma}$ term.

6. The proposed scheme

The proposed algorithm use Fourier magnitudes [8] in place of the original images. As mentioned above, due to the $\alpha-\sigma$ term, the modulus-based FMT descriptors are variant to scale, which contribute to the determine of scale factor. Note that in Eq. (3) the Fourier magnitude of a geometric transformation corresponds to a combination of rotation with angle β and uniform scaling with scale factor $1/\alpha$, without translation, so the AFMT method can be readily used to find a set of appropriate transformation parameters. We employ the AFMT method in form of the Cartesian coordinates instead of the cross-correlation method to circumvent the limitations and problems, such as zero-padding and circularity of cross-correlation, caused by invariant image descriptors in log-polar coordinate space. Especially, the computation cost caused by interpolation during the conversion from Cartesian to log-polar coordinates can be removed by applying the direct approximation in the Cartesian coordinates.

6.1. Approximation

Performing a polar to rectangular change of variable in (6) can obtain another expression of the AFMT. Hence the AFMT is expressed according to the Cartesian coordinates of f [12]:

$$M_{f\sigma}(k,v) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y).$$
$$(x+iy)^{-k} (x^2+y^2)^{\frac{k-2+\sigma-iv}{2}}\, dx \quad (9)$$

In this case, no re-sampling of the discrete image is necessary and the AFMT can be estimated directly from the rectangular grid. Since f vanishes outside a bounded domain of the 2D plane, $f(p, q)$ is considered as a discrete image defined on a rectangular grid with finite spatial extent. The Cartesian AFMT approximation $\hat{M}_{f_\sigma}^{C}$ is computed by using sums in place of integrals:

$$\hat{M}_{f_\sigma}^{C} = \frac{1}{2\pi} \sum_{q=Q\min}^{Q\max} \sum_{p=P\min}^{P\max} f(p,q)$$
$$(p+iq)^{-k} (p^2+q^2)^{\frac{k-2+\sigma-iv}{2}} \quad (10)$$

The coordinates p and q correspond to a pixel position from the image centroid. Pmin, Pmax, Qmin, and Qmax indicate the coordinates, with respect to the image centroid, of the smallest rectangle that fully contains the object. The discrete and finite-extent AFMT $\hat{M}_{f_\sigma}^{C}(k, v)$ is computed for $k \in [-K,K]$ and $v \in [-V, V]$, with a sampling step value over axis v set to 1.

6.2. Implementation

Eq. (8) shows that the AFMT converts a similarity transformation in the original domain into a complex multiplication in the Fourier-Mellin domain. The logarithmic magnitude and phase representation in the Fourier-Mellin domain is written below

$$\ln A_{g_\sigma}(k,v) = -t\sigma + \ln A_{f_\sigma(k,v)}$$

$$\ln g_\sigma(k,v) = vt + k\beta + \ln f_{\sigma(k,v)} \quad (11)$$

where $A_{f_\sigma}(k, v)$ and $\Phi_{f_\sigma}(k, v)$ denote the magnitude and phase representation of the Fourier-Mellin domain for f_σ, respectively; $A_{g_\sigma}(k, v)$ and $\Phi_{g_\sigma}(k, v)$ the magnitude and phase representation for g_σ; and $t = \ln \alpha$. The proposed descriptor is the AFMT of the Fourier magnitude, which removes the effects of translation and rotation, and readily determines the scaling factor, and further, the rotation factor, in terms of the modulus of the Fourier-Mellin transform. Once the optimal rotation and scaling parameters are determined, one of the original images is appropriately scaled and rotated, and the images are cross-correlated to find the optimal translation parameters. The algorithm is written as follows:

1. Calculate the corresponding Fourier magnitude F and G of an image pair f and g, respectively.
2. Calculate the AFMT of F and G by using the Cartesian AFMT approximation given by eq. (10), and obtain $\hat{M}_{f_\sigma}^{C}(k,v)$ and $\hat{M}_{g_\sigma}^{C}(k, v)$, respectively. Following the recommendation of Goh [19], σ is set to 0.5.
3.According to (11), calculate

$$t = \frac{\sum_{v=-V}^{V} \sum_{k=-K}^{K} \frac{\ln \hat{A}_{f\sigma}^{C}(k,v) - \ln \hat{A}_{g\sigma}^{C}(k,v)}{k}}{\sigma(2K-1)(2V-1)} \quad (12)$$

4. With the value t, calculate

$$\beta = \frac{\sum_{v=-V}^{V} \sum_{k=-K}^{K} \frac{\ln \phi_{g\sigma}^{C}(k,v) - \ln \phi_{f\sigma}^{C}(k,v) - vt}{k}}{(2K-1)(2V-1)}$$

$$\alpha = e^{-t} \tag{13}$$

5. One of the image pair is appropriately scaled and rotated with the parameter α and β, and the images are cross-correlated to find the optimal translation parameters (Δx, Δy).

6.3 Syntax Analysis

Characters after recognition are one by one connected in a chain of characters which represents a caption (vehicle identification number). To make a decision if the recognized caption is a license plate a simple syntax analysis is performed. While doing character recognition, computation becomes speedier as only fixed number and sequence of characters are used in Indian number plates, in lieu of identifying the different states to which the car belongs [16]. Each caption is compared with each predefined syntax pattern which represents a proper vehicle identification number. Those patterns describe length of the proper license plate as well as order of character types (digit, letter or alphanumeric character).

7. Experimental Results

To evaluate effectiveness of the method proposed in this paper the overall experiment was divided into the following tests:

T1. Test of effectiveness of Analytical Fourier Transform as compared to original Fourier – Mellin transform for character recognition
In this section, we present the experimental results for testing the performance of the proposed method compared with the methods based on original Fourier-Mellin transform and traditional cross-correlation [4] with added noise and misalignments executed through random scaling and rotational transformations on sample images.

Table 1. Experimental results with simulated misalignment show the root-mean-squared(RMS)errors and computation costs of translation, rotation angle, scale factor for 50 simulations.

Error	Proposed	Traditional	Computation cost ratio
Translation	0.32	0.41	0.861
Rotation	0.54	0.86	0.784
Scaling	0.03	0.09	0.856

Table 2. Experimental results with simulated noise show the root-mean-squared(RMS) errors and computation costs of translation, rotation angle, and scale factor for 50 simulations.

Error	Proposed	Traditional	Computation cost ratio
Translation	0.69	0.72	0.983
Rotation	0.67	0.79	0.876
Scaling	0.07	0.2	0.978

T2. Test of effectiveness of the "signature" technique combined with unique symmetry based localization.
The following section presents the experimental results for testing the performance of the proposed method compared with the traditional signature methods without our proposed symmetry based localization.

Table 3. Experimental results for T2

Error	Proposed	Traditional	Computation cost ratio
	0	0	0.587

T3. Test of effectiveness of license plate character segmentation process based on "peaks to valley" method.
The following section presents experimental results for testing the performance of the process to localize the characters of the license plate which provides segmentation of characters by employing a projection along a horizontal axis of the base-line image to identify positions of the characters, as compared to segmentation algorithms based on Hidden Markov Models [4].

Table 4. Experimental results for T3.

Error	Proposed	Traditional	Computation cost ratio
	0.11	0	0.781

All test were performed on the same data set which consist of 50 various real number plate pictures. The proposed scheme is able to accept any amount of translation and rotation while it accepts scale factors from 0.5 to 2.
Note that computation cost ratio is given by: Computation cost of Proposed method/Computation cost of Traditional method.

8. Future Scope

The method proposed in this paper promises to be very universal and speedy in case of localization and recognition of different license plates under various

environmental and lighting conditions. It has ability to correctly localize license plates located in the picture, in a short time, even if they are dirty or containing small mechanical damages. Further improvement in the scheme can be done by incorporating more basic templates i.e. The reference images for character recognition of alphanumeric characters could be taken in different lighting conditions and environmental conditions. Database in the memory of the system can be updated to include international vehicle license plate schemes.

9. Conclusion

The paper proposed a novel number plate matching method using AFMT. This registration algorithm is able to find transformation parameters to a resolution better than the available discretization of the log-polar parameter space if two images are similar in information content and if the misalignment between them can be modeled with a combination of translation, rotation, and uniform scaling. In addition, the registration algorithm is resistant to white noise. The AFMT algorithm eliminates the re-sampling errors induced in Cartesian to polar Coordinate Conversion in Fourier Mellin Transform.

The main contribution of the method is a solution to the Indian scenario providing a system that is lower in computational costs than traditional methods, and moreover, it directly determines the rotation angle unlike the previous correlation. The symmetry based method of localization using centroid evaluation and the syntax analysis makes the number plate recognition more accurate and efficient as is evident by the analysis. One problem should be solved in the nature is mixed noise which affects the alignment of the two images. When mixed noise is present in the template image and query image, especially the noise belongs to the different types, the AFMT method can not be well executed.

10. References

[1] D. Casasent, D. Psaltis. “Scale invariant optical transform,” Optical Engineering, 15(3):258–261, May/June 1976.Optical Engineering, 15(3):258–261, May/June 1976.

[2] T. Yatagay, K. Choji, H. Saito. “Pattern classification using optical Mellin transform and circular photodiode array, Optical Communication, 38(3):162–165, August 1981.

[3] Y. Sheng, H. H. Arsenault. “Experiments on pattern recognition using invariant Fourier-Mellin descriptors,” J. of the Optical Society of America A, 3(4):771–776, June 1986.

[4] A. E. Grace, M. Spann. “A comparison between Fourier-Mellin descriptors and moment based features for invariant object recognition using neural networks,” Pattern Recognition Letters, 12:635–643, October 1991.

[5] P. E. Zwicke, Z. Kiss. “A new implementation of the Mellin transform and its application to radar classification,” IEEE trans. on Pattern Analysis and Machine Intelligence, 5(2):191–199, March 1983.

[6] F. Ghorbel. “A complete invariant description for gray-level images by the harmonic analysis approach,” *Pattern Recognition Letters*, 15:1043–1051, October 1994.

[7] R. Milanese, M. Cherbuliez. “A rotation-translation-scale-invariant approach to content-based image retrieval,” *J. of Visual Communication and Image Representation*, 10:186–196, June 1999.

[8] R. Bracewell. *The Fourier transform and its applications*, McGraw-Hill Inc., 2nd edition, New York (USA), 1986.

[9] F. Ghorbel. “Towards a unitary formulation for invariant image description: Application to image coding,” *Annals o Telecommunication*, 53(5/6): 242–260, May/June 1998.

[10] P. E. Anuta. “Spatial registration of multispectral and multitemporal digital imagery using fast Fourier transform techniques,” *IEEE Trans. on Geo. Elec.*, 8:353–368, 1970.

[11] D. Casasent, D. Psaltis. “Position, rotation, and scale invariant optical correlation,” *Applied Optics*, 15:1795–1799, 1976.

[12] D. Casasent, D. Psaltis. “Space-bandwidth product and accuracy of the optical Mellin transform,” *Applied Optics*,16:1472, 1977.

[13] D. Casasent, D. Psaltis. “Deformation invariant, spacevariant optical pattern recognition,” In Wolf E (Ed): Progress in Optics, North-Holland Publishing Co, Amsterdam, 290–356, 1978.

[14] A. S. Jensen, L. Lindvold, E. Rasmussen. “Transformation of image positions, rotations, and sizes into shift parameters,” *Applied Optics*, 26:1775–1781, 1987.

[15] S. Derrode, F. Ghorbel. “Robust and efficient Fourier-Mellin transform approximations for graylevel image reconstruction and complete invariant description,” *Computer Vision and Image Understanding*, 83(1):57–78, 2001.

[16] www.tn.gov.in/sta/regmark.htm

[17] www.morth.nic.in

[18] Canny, J, A Computational Approach to Edge Detection , IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, pp. 679-698, 1986.

[19] S. Goh. “The Mellin transformation: Theory and digital filter implementation,” Ph.D. dissertation, Purdue University, West Lafayette, I.N., 1985.

Middleware-Based Kalman Filter Design for a Driver's Aid System

Wenwei Hou[1], Cecil Bruce-Boye[2], Dmitry A. Kazakov[3] and Youling Zhou[4]

Abstract— **In this paper we present an implementation of the Kalman filter based on a middleware. This middleware LabMap is applied as a data acquisition and distribution layer in industrial automation area, serving as a software bus. We use its virtual handles to implement the algorithm of the Kalman filter directly at the level of the software bus. The advantage of this approach is that neither of the system components needs to be changes as the implementation is enclosed in the middleware. We show that the Kalman filter is useful to remove the speed noise in a driver's aid system.**

Note to Practitioners— **This paper was motivated by the problem of the speed noise in the driver's aid system ErgoDrive Professional for the visual instruction of test drivers on chassis dynamometers. The speed data from the hardware has noise. This leads to trembling of the speed cursor in the driver's aid, which makes it is difficult for the driver to drive the test profile. In this work we applied the Kalman filter to remove the speed noise. Our experiments showed that we always were able to find suitable parameters for the Kalman filter for sampling rates up to 100ms. For larger sampling time intervals performance degrades. In future research we will try to solve this problem with a predictive filter.**

Index Terms— **Middleware, Software bus, LabMap, Kalman filter**

I. Introduction

In a distributed computing system, middleware is defined as a software layer between software components used to reduce coupling.

With the growth of network-based applications, middleware technologies become more and more important. They cover a wide range of software systems, including distributed objects and components, message-oriented communication, and mobile application support.

The role of middleware is to make application development easier, by providing common programming framework, by abstracting heterogeneous and distributed nature of the underlying hardware and software components, and by hiding lower-level programming details.

LabMap [1, 2, 3, 4] represents an example of such middleware used in the process automation and data acquisition area. It provides an abstraction of the application level from the hardware specifics by decoupling the hardware interface modules from the application level. It also allows a smooth integration of numerous components with their variety of software and hardware protocol, supporting a component based software design. Another important advantage of this particular middleware is that it has computable data channels, so-called virtual handle, so that a definite data processing can implement at the level of middleware without modification of other software components.

ErgoDrive Professional [5, 6] is a driver's aid system used to in chassis dynamometer automation solutions and test environments. It is being deployed in many application areas such as exhaust emission certification, brake force measurements, vehicle cooling testing, fuel consumption testing, gear box optimizations.

The Kalman filter [7] is a set of mathematical equations that provides an efficient computational (recursive) means to estimate the state of a process, in a way that minimizes the mean of the squared error. The filter is very powerful in several aspects.

In this paper we will use a new tool of the middleware LabMap virtual handle, to implement the Kalman filter and show it is useful to remove the noise of the speed in the driver's aid system.

II. The Middleware Architecture

As a distributed middleware LabMap is operating as a software bus. It provides a high level interface for industrial communication, control, measurement, and logging. It allows asynchronous viewing and modification of the process variables from a potentially unlimited number of applications running on the same or different LAN/WAN network hosts. The middleware has an open architecture supporting smooth integration of new hardware and software components running on different bus architectures. It has integrated support of measurement units and time stamping of the system variables.

This software bus offers an abstraction of the application level from the hardware specific and decoupling the hardware interface modules from the application level. In other words there are two levels of abstraction supported by the bus system:

A. The application interface

B. The hardware driver interface

The application abstraction interface exposes LabMap as a set of variables (registers). Each register has a type, value, timestamp and I/O direction. The four basic I/O requests are supported for each register:

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 43–48.

1) Get

The value of a register can be read in a safe way that warranties consistency of the read value bits and the timestamp. The application is relieved from the burden of locking the value in presence of other tasks accessing it concurrently.

2) Set

The value of a register can be set in a safe way. No I/O initiated by this request.

3) Request (for output registers only)

A new value of a register can be requested from the underlying hardware. The application needs not to know which actions are necessary to request the new value. It is the responsibility of the driver. The request is asynchronous and the application is not blocked until I/O completion. However it may enter a non-busy waiting for the I/O completion.

4) Send (for input registers only)

The actual value of a register can be sent to the underlying hardware. Like in the case of request the application is unaware of the actions the hardware undertakes upon send. The application may enter a non-busy waiting for the I/O completion.

The hardware abstraction interface of the software bus allows an easy integration of new hardware devices and communication protocols into the system. Each class of hardware devices or communication protocol supported by the software bus is represented by a hardware interface. The hardware interface is responsible for implementation of the basic I/O operations upon the registers assigned to it. The software bus uses a messages mechanism for interacting with the hardware interfaces. A hardware interface is developed as dynamic-link library which code can be maintained independently from the middleware core.

C. Computable registers

Computable registers are represented in the middleware LabMap as a virtual hardware. The corresponding interface is named LabVirt. In this paper we use this functionality to implement a Kalman algorithm solely in the middleware without modification of the driver's aid component. The computable registers interface gives an ability to calculate registers out of other register values. The way a register is evaluated is determined by its configuration, which basically a program specified in the language of LabVirt. It is a compiled language. The code is generated for the virtual machine integrated into the computable registers interface implementation.

For example if we wanted to add values of two registers and store the result into a third register, we could define the virtual register like:

($100 + $200) -> 300

III. The Algorithm of the Discrete Kalman Filter

The Kalman filter is a set of mathematical equations that provides efficient computational means to estimate the state of a process, in a way that minimizes the mean of the squared error. The filter is very powerful in several aspects: it supports estimations of past, present, and even future states, and it can do so even when the precise nature of the modeled system is unknown [8, 9, 10].

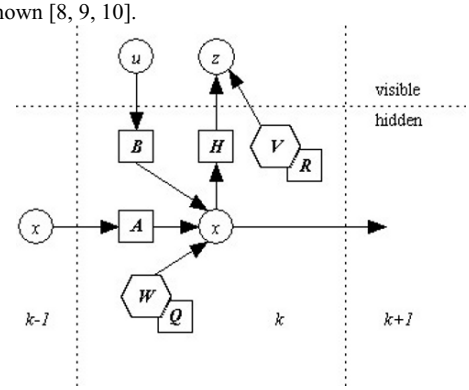


FIG.1 The model of the Kalman filter

A. The Process To Be Estimated

The Kalman filter addresses the general problem of trying to estimate the state of a discrete-time controlled process that is governed by the linear stochastic difference equation

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \tag{1}$$

with a measurement that is

$$z_k = Hx_k + v_k \tag{2}$$

The random variables w_k and v_k represent the process and measurement noise (respectively). They are assumed to be independent (of each other), white, and with normal probability distributions

$$p(w) \sim N(0, Q), \tag{3}$$

$$p(v) \sim N(0, R). \tag{4}$$

In practice, the process noise covariance Q and measurement noise covariance R matrices might change with each time step or measurement, however here we assume they are constant.

The $n \times n$ matrix A in the difference equation (1) relates the state at the previous time step $k-1$ to the state at the current step k, in the absence of either a driving function or process noise. The matrix B relates the optional control input $u \in R^l$ to the state x. The $m \times n$ matrix H in the measurement equation (2) relates the state to the measurement z_k. In practice A, B and H might change with each time step or measurement, but here we assume it is constant.

B. The Discrete Kalman Filter Algorithm

The Kalman filter estimates a process by using a form of feedback control: the filter estimates the process state at some time and then obtains feedback in the form of (noisy) measurements. As such, the equations for the Kalman filter

fall into two groups: time update equations and measurement update equations. The time update equations are responsible for projecting forward (in time) the current state and error covariance estimates to obtain the a priori estimates for the next time step. Indeed the final estimation algorithm resembles that of a predictor-corrector algorithm for solving numerical problems as shown below in Fig. 2.

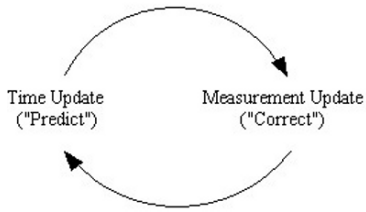


FIG.2 The ongoing discrete Kalman filter cycle

The specific equations for the time and measurement updates are presented below.

Discrete Kalman filter time update equations:

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1} \tag{5}$$

$$P_k^- = AP_{k-1}A^T + Q \tag{6}$$

Discrete Kalman filter measurement update equations:

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \tag{7}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \tag{8}$$

$$P_k = (I - K_k H)P_k^- \tag{9}$$

IV. Application of the Kalman filter for driver's aid system

A. System Structure

The system structure is shown on Fig.3. On the computer LabMap runs with registers configured for a DAC and ADC modules used to set a defined voltage over the amplifier to the motor and get the voltage signal generated by the tacho-machine. The modules are connected to the middleware using the ModBus protocol. The hardware interface responsible for its implementation in the middleware is LabModBus. The corresponding variables are the registers served in the middleware by this interface.

There are acceleration and brake pedals used for the user input. The values generated by the pedals according to the force applied are acquired by an application and stored into the software bus using the hardware interface called LabUser. This interface allows a direct control over the value of the corresponding variables without having any hardware behind them.

The variables controlling the driver's aid are served by the LabFLG hardware interface. The Ergo Drive driver's aid software is treated as a hardware is this setup.

The variables implementing the Kalman filter are served by the LabVirt interface.

The variables used for setting the filter parameters are LabUser registers.

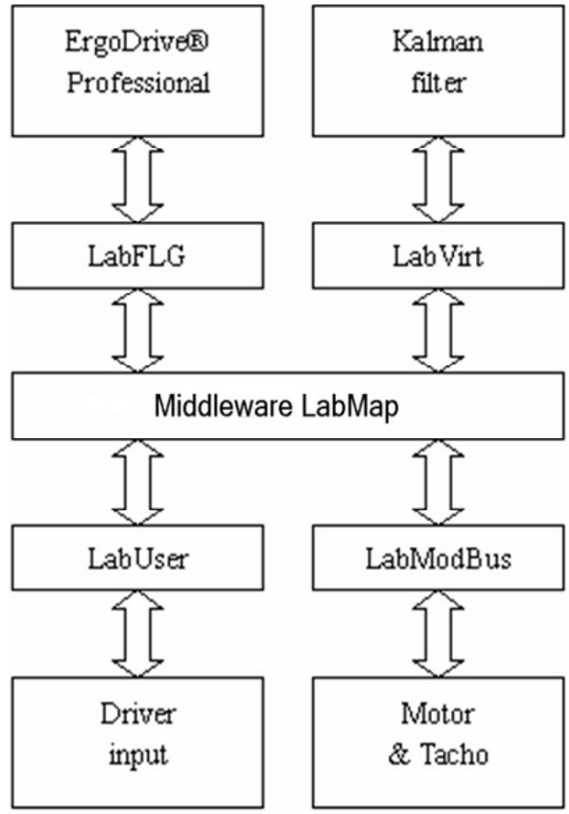


FIG.3 System structure

Due to the distributed nature of LabMap we were able to test and use different setups of our system in which some of the system components run on physically different computers.

B. The Process Model

We try to remove the noise from the speed of the motor that we get from indirectly the tacho-machine. Let us assume that we give a constant voltage to the motor, so the speed of the motor is also a constant value, so A=1.There is no control input so u =0. So

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} = x_{k-1} + w_{k-1} \tag{10}$$

The tacho-machine can generate voltage according to the speed of the motor. The relationship between voltage and speed is 0.52V per 1000rpm that is 0.0312V*s. So we can calculate the frequency from the voltage. On the other hand we can get the speed of motor by multiplying frequency with the girth of wheel. The input of the filter is that speed so H =1. So

$$z_k = Hx_k + v_k = x_k + v_k \tag{11}$$

Our time update equations are

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1} = \hat{x}_{k-1} \tag{12}$$

$$P_k^- = AP_{k-1}A^T + Q = P_{k-1} + Q \quad (13)$$

And our measurement update equations are

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} = P_k^- (P_k^- + R)^{-1} \quad (14)$$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) = \hat{x}_k^- + K_k(z_k - \hat{x}_k^-) \quad (15)$$

$$P_k = (I - K_k H)P_k^- = P_k^- - K_k P_k^- \quad (16)$$

Let us substitute (13) into (14) so we have

$$K_k = (P_{k-1} + Q)(P_{k-1} + Q + R)^{-1} \quad (17)$$

Let us substitute (12) and (17) into (15) so we have

$$\hat{x}_k = \hat{x}_{k-1} + (P_{k-1} + Q)(P_{k-1} + Q + R)^{-1}(z_k - \hat{x}_{k-1}) \quad (18)$$

Let us substitute (13) and (17) into (16) so we have

$$P_k = P_{k-1} + Q - (P_{k-1} + Q)^2 (P_{k-1} + Q + R)^{-1} \quad (19)$$

So the key to implement this Kalman filter by the middleware is to define two virtual-Registers with the filter algorithm (18) and (19). The software bus LabMap configuration is presented in the Table 1.

TABLE I
LABMAP REGISTER CONFIGURATION

Register No.	Configuration		Description
10000	>User> [::]!O:R:0:0:-1:-1:: Y_Koordinate		The value got from the joy stick.
10002	>Virt> [V:V:V]!O:R:0:0:-1:-1:: joy value to motor voltage 10000- >[(value*0.00030517578 125-9.99969482421875 max 0)*1[V]->710]		Calculate the motor voltage from the value of joy stick [2]
710	>ModBus> [V:V*0.000305185:V]!I: R:0:0:-1:-1:: Motor voltage [Coupler1 type=i addr=0 Len=16]		ModBus input register to set a defined voltage (+/- 10V) to the Wago analogue output module on output channel 0 -> motor voltage.
720	ModBus> [V:0.000305185*V+0.043 3363:V] !O:R:0:10:-1:-1:: tacho voltage [Coupler1 type=i addr=0 Len=16]		ModBus output register to get the noised voltage from the Wago analogue input module on input channel 0 <- tacho voltage.
721	>Virt> [km/h:km/h:km/h]!O:R:0: 0:-1:-1:: convert filter 720- >[$720/0.0312[V*s]*0.29 9[m]*3600[s/h]/1000[m/k m]=>722]		Convert the voltage to the speed
722	>User> [km/h:km/h:km/h]!O:R:0: 0:-1:-1:: orignal speed	z_k	The input of the filter
730	>User> [km/h:km/h:km/h]!I:R:0:0 :-1:-1:: speed_kalman	$\hat{x}_k$	User register for the posteriori estimate speed of the motor. Before testing the filter the register must be set with a defined value.g
740	>User> [km²/h²: km²/h²: km²/h²] !I:R:0:0:-1:-1:: Covariance	P_k	User register for the posteriori estimate error covariance. Before testing the filter the register must be set with a defined value.
750	>User> [km²/h²: km²/h²: km²/h²] !I:R:0:0:-1:-1:: Q	Q	User register for process noise covariance. It can be modified at runtime.
760	>User> [km²/h²: km²/h²: km²/h²] !I:R:0:0:-1:-1:: R	R	User register for measurement noise covariance. It can be modified at runtime.
770	>Virt> [km²/h²: km²/h²: km²/h²] !O:R:0:0:-1:-1:: covariance_filter 722->[($740+$750)*(1- 1/(1+$760/($740+$750))) ->740]		Virtual output register with filter algorithmic of the posteriori estimate error covariance.
780	>Virt> [km/h:km/h:km/h] !O:R:0:0:-1:-1:: speed_filter 740->[$730+($722- $730)/(1+$760/(delay(val ue)+$750))->730]		Virtual output register with filter algorithmic of the posteriori estimate speed of the motor

V. RESULTS

To use the filter firstly we must set suitable value to Q and R. Their values have great influence on the performance of the filter. So we always have to find a relatively better parameter, which can keep the output smooth and do not have too much time delay.

To use the filter we also must set initial value to the speed register and the covariance register. From the Kalman filter theory, we know that the alternative choice is not critical. We could choose almost any $P_0 \neq 0$ and $\hat{x}_0$, the filter would eventually converge.

We try different values of Q and R to test the influence of them on the output signal. We can conclude that in the case of keeping R as a constant when Q becomes smaller, the output of the filter will become smoother but at the same time the time delay will become bigger.

On the other hand, in the case of keeping Q as a constant when R becomes bigger, the output of the filter will become smoother but at the same time the time delay will become bigger.

Finally we find that if sampling interval is 10ms, then Q=0.000001 km²/h², R=0.001 km²/h² would be a relatively better parameter.

Fig. 4 represents the input of the filter (blue line) and the output of filter (red line). From the figure we can see that the output signal has a delay around 150ms.

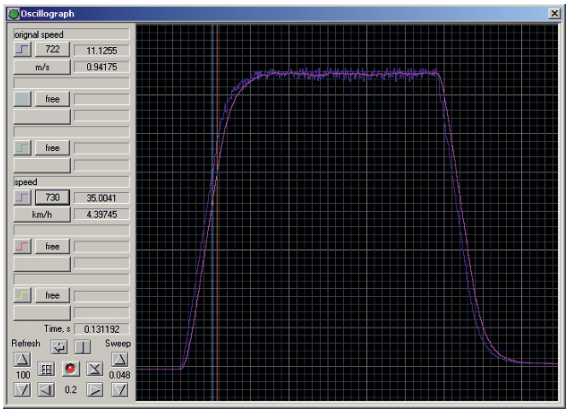


FIG.4 Q=0.000001 km²/h², R=0.001 km²/h²

The figure is obtained from the oscilloscope tool integrated in the LabMap.

After some more experiment, we find when we change the sample time of the input signal we always can find a relatively better parameter of the filter.

TABLE II
DIFFERENT PARAMETER FOR DIFFERENT SAMPLING INTERVAL

Sample time (ms)	Q (km²/h²)	R (km²/h²)	Time delay (ms)
10	0.000001	0.001	150
20	0.00001	0.001	100
30	0.00001	0.001	100
50	0.00001	0.0001	50
70	0.00001	0.0001	70
100	0.00001	0.0001	100

This kind of Kalman filter has been used in a driver's aid system. It is useful to remove the noise of the speed signal from the chassis dynamometer so that the driver can see a smooth time-speed line on the screen. Fig. 5 represents the driver's aid display when driven without the filter.

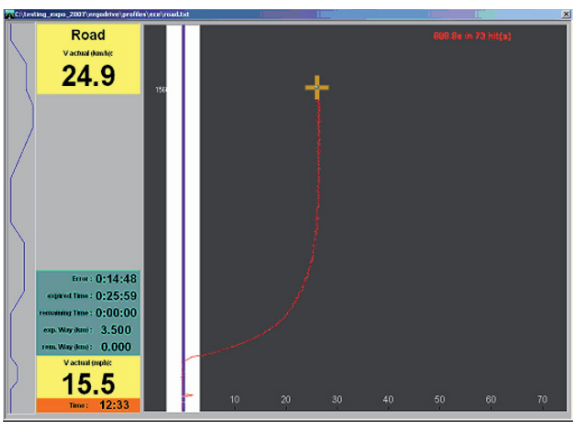


FIG.5 Driver's aid behavior without filtering

Fig. 6. represents the driver's aid performance with the Kalman filter we have designed.

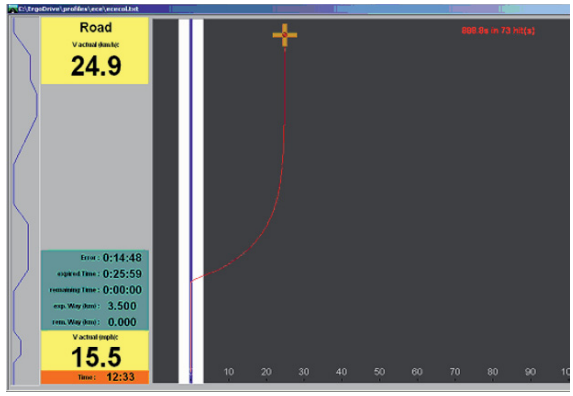


FIG.6 Driver's aid behavior with the Kalman filter

VI. CONCLUSION

We have observed that this software bus with its computable data channels is suitable for simple digital signal processing as required in engineering tasks. Some problems like on-line signal filtering can be designed, tested and deployed without modification of the legacy components. For an implementation of the algorithms such as the Kalman filter in the software bus, we need to define proper input and output registers and describe the algorithm in the language of the computable registers.

REFERENCES

[1] C. Bruce-Boye and D. Kazakov, "Distributed data acquisition and control via software bus," Proceedings CSMITA'04, pp. 153–156, Sep 2004.

[2] C. Bruce-Boye, D. Kazakov, Quality of Uni- and Multicast Services in a Middleware. LabMap Study Case," Conference CIS²E 06 (International Joint Conference on Computer, Information and System, Science and Engineering") 2006, IEEE, 4-14 December 2006

[3] C. Bruce-Boye, D. Kazakov; Rüdiger zum Beck, "An approach to distributed remote control based on middleware technology, MATLAB/Simulink-LabMap/LabNet framework", Conference CIS²E (International Joint Conference on Computer, Information and System, Science and Engineering") 2005, IEEE, 10-20 December 2005.

[4] C. Bruce-Boye, D. Kazakov, "Distributed data acquisition and control via software bus", International Industrial Ethernet Development High Level Forum 2004 (IEHF 2004) in Peking, Automation Panorama No. 5

[5] R. Müller, D. Kazakov, A. Fechner, A. Wilde, „A smooth ride", Testing technology International", pp.82-84, Nov. 2000

[6] "GPS Recorder", Testing Technology International, p.95, November 2004

[7] Kalman, R. E., "A new approach to linear filtering and prediction problems," Transaction of the ASME—Journal of Basic Engineering, March 1960, pp. 35-45

[8] Brown, R. G. and P. Y. C. Hwang, "Introduction to Random Signals and Applied Kalman Filtering", Second Edition, John Wiley & Sons, Inc, 1992

[9] Grewal, Mohinder S., and Angus P. Andrews. "Kalman Filtering Theory and Practice". Upper Saddle River, NJ USA, Prentice Hall, 1993

[10] Sorenson, H. W. "Least-Squares estimation: from Gauss to Kalman," IEEE Spectrum, vol. 7, pp. 63-68, July 1970.

[1]**Wenwei Hou**, East China University of Science and Technology, 130 Meilong Road, Shanghai, 200237, P.R.China

[2]**Cecil Bruce-Boye**, University of applied science, 3 Stephensonstrasse Luebeck, 23562 Germany
[3]**Dmitry A. Kazakov,** cbb software GmbH, 1 Charlottenstrasse Luebeck, 23560 Germany
[4]**Youling Zhou,** East China University of Science and Technology, 130 Meilong Road, Shanghai, 200237, P.R.China

ACKNOWLEDGES

We should give special thanks to Luebeck University of Applied Sciences and East China University of Science and Technology. They launch the exchange student program supported by DAAD, so that we can work together and write this paper.

Improving Neural Network Performances – Training with Negative Examples

COSMIN CERNĂZANU-GLĂVAN, ŞTEFAN HOLBAN
Computer Science Department
"Politehnica" University of Timisoara
Pta.Victoriei no.2, RO-300006 Timisoara
ROMANIA
cosmin.cernazanu@ac.upt.ro

***Abstract*-The paper describes a new concept for training the neural networks, which can be concretized by using a new modified classical architecture. The idea considered to be the basis of this new architecture was taken over from learning machine, namely that for defining a concept, we need both negative and positive examples. Neural networks are models that are trained only with positive examples and allow the recognition of new examples using the learning machine. Training neural networks with negative examples aims at preventing the development of some specific features of these examples and at obtaining a better recognition of the positive examples. The architecture developed through this method is generic and can be applied to any type of neural network. For simplicity and for the need of obtaining immediate results, a multilayer perceptron was chosen for testing. The results achieved with the help of this network are encouraging and they open new possibilities of study for the future.**

I. What Is A Neural Network?

An artificial neural network is a model that emulates the biologic neural network. An artificial neural network is made up of thousands of artificial neurons; elements of non-linear processing that operate in parallel. [1]

The main characteristics of the neural networks are the same with those of the human brain, that is:

- capacity of learning
- capacity of generalizing

If trained adequately, the artificial neural networks would be capable of providing correct answers even for the set-entries different from those they have already been used to, as long as they do not differ too much. This generalization is made automatically as a result of their structure and not as a result of human intelligence which is included in a program as in the case of the expert systems.

II. Improving Neural Network Performances – Training With Negative Examples

A. The Problem

Let's presume that we need a neural network for an off-road vehicle specialized in collecting special rocks from hard to reach places (volcanoes, oceans, other planets, etc.). For this purpose, we trained the neural network to recognize certain types of rocks (we necessarily need samples from these rocks) that we hope to adequately resemble the ones present in the places mentioned above. One of the main features of the neural network, namely the capacity of generalizing, enables the identification of many common rocks as being special rocks, and this thing happens frequently if we have a small training set. For this reason, it is desirable that we training the neural network with some features that do not belong to the special rocks. Through this operation, we will facilitate the neural network training both by providing new examples for training the neural network and by emphasizing the features of the elements that should be recognized. In other words, if the network is not 100% sure whether the rock is special or not and it does not recognize the class to which the rock belongs, it may, in some cases, be 100% sure that the respective rock is not special.

B. The Idea

The use of negative examples for training the neural networks is an old-established technique in the field. Therefore, the training set must include many negative examples as well, otherwise the neural network will provide only optimistic answers and it will output "yes" all the times.

There existed some attempts [18][19][20] to generate some virtual examples especially for a better training of the network.

To conclude, it should be mentioned that we should step forward and give the negative examples the same importance we give to the positive ones.

The idea comes from machine learning field. [10][11][12][13][14] In general, machine learning involves adaptive mechanisms that enable computers to learn from experience, from examples and by analogy [2][5]. As a result, the performance of computers increased over time.

As we can see in [3] and in Fig. 1, for learning the concept of "an arch", 4 examples have been suggested: 2 negative examples and 2 positive examples. Learning from examples is

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 49–53.

also called inductive learning.[7][8][9] The computer can be made to learn the concept of an arch from teacher-supplied positive examples and negative examples. The examples are processed one by one and the computer modifies the definition (named "current hypothesis") to sustain all the processed examples. We may add that the definition of an object may include both features that belong to the object and features that do not belong to the object. As a consequence, the recognition of both types of features and their combination at the level of the neural networks is necessary, so that only the objects with necessary features (and not the unnecessary ones) be selected.

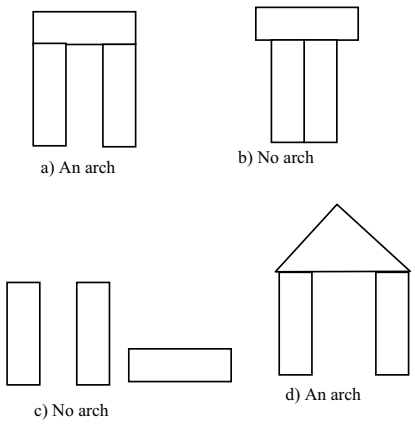


Figure 1. Definition of the arch

C. Neural Network – Multilayer Perceptron

As architecture, the multilayer perceptron[4] is represented as we can see in Fig. 2.

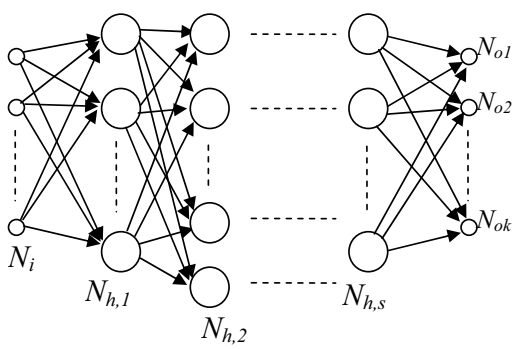


Figure 2. Multilayer perceptron architecture

The first layer N_i is the input layer. The neurons on this layer lack the activating function, allowing any value received as input to pass. The following s layers ($N_{h,1}$, .., $N_{h,s}$) are named hidden layers. Each hidden layer ($N_{h,l}$), receives input values only from the previous layer ($N_{h,l-1}$), its outputs representing inputs for the next layer ($N_{h,l+1}$). No connection is allowed inside the layer.

The outputs of the last hidden layer $N_{h,s}$ represent inputs for the output layer N_o. We must add that connections of the type *skip layer* can be allowed (the neurons can receive as inputs the outputs of other neurons that are not placed on a previous layer in the immediate neighborhood). In general, the direct connections between the input and the output layers are very useful.

The number of neurons on the hidden layers can differ for each layer. Each neuron has an activating function F_i that works on inputs and bias.

D. How the Idea is Applied to Neural Networks

The change produced to the neural network consists in adding a number of t neurons on the output layer (N_{ob1} ..N_{obt}). (Figure 3) These neurons are named block neurons and are responsible for blocking all the negative examples.

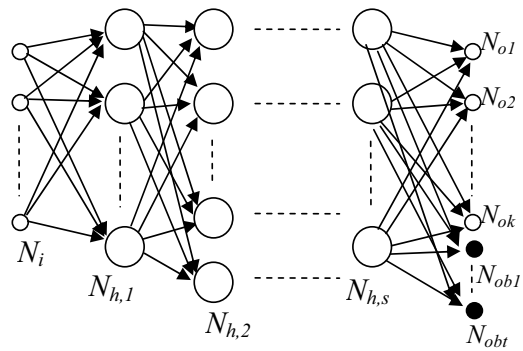


Figure 3. Multilayer perceptron architecture with k output neurons for recognition and t output neurons for blocking neural patterns

When these neurons are active, we can say that the pattern placed at the input of the neural network is not a pattern that should be recognised by the neural network.[15][16][17] In other words, the block neurons are responsible for recognizing the features that do not define the class of objects that must be recognised.

Example : if we train a neural network for recognizing the letters a,o,c,u,e, we can take as negative examples the letters d,t,b,g,p, and the feature that does not belong to the first class would be the "tail" that belongs to the letters from the second class.

III. Implementing And Testing A Neural Network Trained Both With Positive And Negative Examples

It is difficult to make a comparison with other results as the present idea is new and it implies an architecture completely different from that of the already existing neural networks. Consequently, it is difficult, if not impossible to find previous results in order to make a comparison between the block neurons architecture and a current architecture.

For this reason, we have carried out an implementation by using the Neuro Shell program in which we have chosen a feed forward neural network with one hidden layer. As set for training, we used the database NIST 19 which contains over 800 000 letters [6].

For testing the performance, we will implement 4 different architectures for each neural network.

A. First Network

This is a classical usage of a neural network. We will try training a neural network for 5 letters a, c, e, o, u, each letter having 100 patterns. A letter is represented as a 32x32 dots matrix. The output will be either 0 or 1.

The neural network is made up of 3 layers (input layer, hidden layer, output layer) having the following configuration:

-the input layer – 1024 neurons, the scalar function being the linear function f(x)=x. The inputs can be either 1 or 0.

-the hidden layer – 532 neurons, the activating function being of the logistic type f(x)=1/(1+exp(-x))

-the output layer – 5 neurons, the activating function being of the logistic type f(x)=1/(1+exp(-x))

B. Second Network

In addition to the first network, we introduced negative examples. Thus, we will enter as negative patterns the following letters: b, d, g, p, t, each letter having 30 patterns.

The neural network is made up of 3 layers having the following configuration:

-the input layer – 1024 neurons, the scalar function being the linear function f(x)=x. The inputs can be either 1 or 0.

-the hidden layer – 532 neurons, the activating function being of the logistic type f(x)=1/(1+exp(-x))

-the output layer – 6 neurons, the activating function being of the logistic type f(x)=1/(1+exp(-x)) . Five from six neurons are used to recognize one of the letters: a, c, e, o, u. The last neuron is the block neuron and it is active when b, d, g, p, t, patterns are trained.

C. Third Network

For this network we used the same patterns as for the second network. The difference consists in the number of hidden layers. Thus, we will have two hidden layers which contain 269 neurons each.

D. Fourth Network

It resembles the second network, but we used 2 block neurons instead of 1.

All networks were trained for 500 epochs. We will try the recognition/non-recognition from a set of 260 letters (10 patterns from each letter of the alphabet) for each of the 4 networks.

E. The Results

The results obtained are presented in TABLE I and graphicaly in Figure 4.

TABLE I
EXPERIMENT RESULTS FOR ALL NETWORKS

Network Number	a,c,e,o,u Recognition	All letters recognition
Network 1	62 %	30.3 %
Network 2	48 %	68.8 %
Network 3	40 %	69.6 %
Network 4	50 %	70,7 %

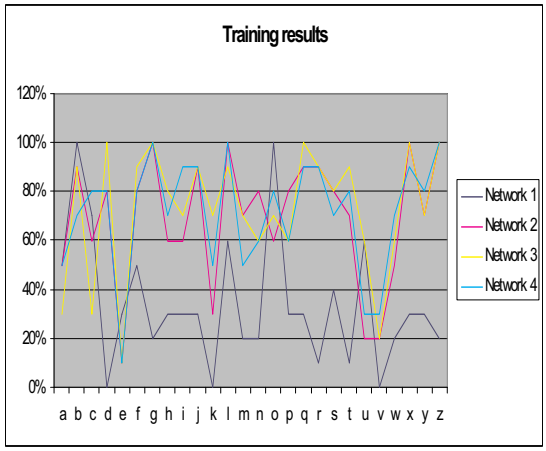


Figure 4. Overview of the training results

As a first remark, we can say that the letter recognition rate decreases proportionally to the number of negative templates added, which is normal in the conditions in which the more templates for recognition we add, the more the recognition rate will decrease.

The second remark is that the recognition percent for the letters foreign to the training doubled in comparison to the first neural network used.

In what follows we will present details concerning each type of implementation. Each column has a specific signification:

-first column – the letter that has to be recognised

- a,c,e,o,u columns – the percent that represents the number of templates recognized as being the corresponding letter for each column

- unknown – the letter was not recognized by the network, or the system marked the letter as unknonown letter

For the first and the second neural networks we can compare the data from TABLE II and TABLE III.

Considering TABLE II and TABLE III, it can be easily noticed that by using the "training neural networks with negative examples" technique, we succeeded to perform a 100% recognition for some letters that do not belong to the training set and that were not previously known by the network. Therefore, the letters z, x, l have been fully recognized as being unknown to the network as compared to the percentages of 20,30% and 60% (percentages obtained by using the same letters, but with the help of the first network). We can add that the recognition rate for the entire set of letters

doubled in comparison to the other ones, recording an increase from 30,3% to 68,8%.

TABLE II
EXPERIMENT RESULTS FOR NETWORK 1

Letter	a	c	e	o	u	unknown
a	50%	0%	0%	10%	10%	30%
b	20%	10%	20%	20%	20%	10%
c	0%	70%	0%	0%	0%	30%
d	0%	0%	0%	0%	100%	0%
e	10%	40%	30%	0%	0%	20%
f	0%	10%	0%	20%	20%	50%
g	30%	0%	10%	10%	30%	20%
h	0%	0%	10%	0%	60%	30%
i	0%	40%	0%	10%	20%	30%
j	20%	10%	10%	0%	30%	30%
k	10%	10%	60%	0%	20%	0%
l	0%	0%	0%	10%	30%	60%
m	10%	0%	10%	20%	40%	20%
n	0%	0%	20%	50%	10%	20%
o	0%	0%	0%	100%	0%	0%
p	10%	0%	20%	20%	20%	30%
q	20%	0%	30%	0%	20%	30%
r	10%	0%	50%	30%	0%	10%
s	40%	20%	0%	0%	0%	40%
t	20%	0%	20%	0%	50%	10%
u	30%	0%	0%	0%	60%	10%
v	0%	0%	0%	0%	100%	0%
w	10%	0%	20%	0%	50%	20%
x	20%	10%	20%	0%	20%	30%
y	0%	0%	0%	0%	70%	30%
z	30%	20%	30%	0%	0%	20%

TABLE III
EXPERIMENT RESULTS FOR NETWORK 2

Letter	a	c	e	o	u	unknown
a	50%	0%	0%	0%	20%	30%
b	10%	0%	0%	0%	0%	90%
c	10%	60%	0%	0%	0%	30%
d	0%	0%	0%	0%	20%	80%
e	10%	50%	10%	0%	0%	30%
f	0%	10%	0%	0%	10%	80%
g	0%	0%	0%	0%	0%	100%
h	0%	0%	0%	0%	40%	60%
i	0%	20%	0%	0%	20%	60%
j	10%	0%	0%	0%	0%	90%
k	0%	10%	20%	0%	40%	30%
l	0%	0%	0%	0%	0%	100%
m	10%	0%	0%	0%	20%	70%
n	0%	0%	0%	20%	0%	80%
o	0%	0%	0%	60%	0%	40%
p	0%	0%	0%	0%	20%	80%
q	0%	0%	0%	0%	10%	90%
r	0%	0%	10%	0%	0%	90%
s	20%	0%	0%	0%	0%	80%
t	10%	0%	10%	0%	10%	70%
u	20%	0%	0%	0%	60%	20%
v	0%	0%	0%	0%	80%	20%
w	0%	0%	0%	0%	50%	50%
x	0%	0%	0%	0%	0%	100%
y	0%	0%	0%	0%	30%	70%
z	0%	0%	0%	0%	0%	100%

For the third and the fourth neural network we can compare the data from TABLE IV and TABLE V. The last two networks provide slightly better results than the second one.

TABLE IV
EXPERIMENT RESULTS FOR NETWORK 3

Letter	a	c	e	o	u	unknown
a	30%	0%	0%	10%	0%	60%
b	0%	0%	0%	0%	10%	90%
c	10%	30%	0%	0%	0%	60%
d	0%	0%	0%	0%	0%	100%
e	0%	50%	10%	0%	0%	40%
f	0%	0%	0%	0%	10%	90%
g	0%	0%	0%	0%	0%	100%
h	0%	0%	0%	0%	20%	80%
i	0%	0%	0%	0%	30%	70%
j	10%	0%	0%	0%	0%	90%
k	0%	10%	0%	0%	20%	70%
l	0%	0%	0%	0%	10%	90%
m	10%	0%	0%	20%	0%	70%
n	0%	0%	0%	40%	0%	60%
o	0%	0%	0%	70%	0%	30%
p	0%	0%	0%	0%	40%	60%
q	0%	0%	0%	0%	0%	100%
r	0%	0%	0%	0%	10%	90%
s	20%	0%	0%	0%	0%	80%
t	0%	0%	10%	0%	0%	90%
u	20%	0%	0%	0%	60%	20%
v	0%	0%	0%	0%	80%	20%
w	0%	0%	0%	0%	40%	60%
x	0%	0%	0%	0%	0%	100%
y	0%	0%	0%	0%	30%	70%
z	0%	0%	0%	0%	0%	100%

TABLE V
EXPERIMENT RESULTS FOR NETWORK 4

Letter	a	c	e	o	u	unknown
a	50%	0%	0%	0%	0%	50%
b	10%	0%	0%	20%	0%	70%
c	0%	80%	0%	0%	0%	20%
d	0%	0%	0%	0%	20%	80%
e	0%	60%	10%	0%	0%	30%
f	0%	10%	0%	0%	10%	80%
g	0%	0%	0%	0%	0%	100%
h	0%	0%	0%	0%	30%	70%
i	0%	10%	0%	0%	0%	90%
j	10%	0%	0%	0%	0%	90%
k	0%	10%	30%	0%	10%	50%
l	0%	0%	0%	0%	0%	100%
m	10%	0%	10%	20%	10%	50%
n	0%	0%	0%	40%	0%	60%
o	0%	0%	0%	80%	0%	20%
p	0%	10%	0%	10%	20%	60%
q	0%	0%	0%	0%	10%	90%
r	0%	0%	10%	0%	0%	90%
s	30%	0%	0%	0%	0%	70%
t	10%	0%	10%	0%	0%	80%
u	20%	0%	0%	0%	30%	50%
v	0%	0%	10%	0%	60%	30%
w	0%	0%	0%	0%	30%	70%
x	0%	10%	0%	0%	0%	90%
y	0%	0%	0%	0%	20%	80%
z	0%	0%	0%	0%	0%	100%

IV. Conclusions

We must add that this network was build only for experimental usage. The results obtained through this experiment are promising and lead to future research concerning this new architecture.

This technique was applied on a set of characters, out of which only 5 of them were already known by the neural network. The rest of the characters had to be identified by the network as being foreign characters. We can state that this technique has a recognition rate (on the level of a set of 26 characters) twice as good as the previous similar recognition techniques used in the field of neural networks.

V. References

[1] Teodorean Gavril, "Artificial Neural Networks", Ed. Cluj-Napoca, 1995

[2] Negnevitsky, M. "Artificial Intelligence: A Guide to Intelligent Systems" (2nd Edition), Addison Wesley, England, 2005.

[3] Bratko I. "PROLOG – Programming for Artificial Intelligence" (2nd Edition) Addison Wesley, England,1993.

[4] Russell S., Norvig P. – "Artificial Intelligence : A Modern Approach" (Second Edition), Prentice Hall,2003.

[5] Luger G. – "Artificial Intelligence :Structures and Strategies for Complex Problem Solving" (Fifth Edition) Addison Wesley, 2005.

[6] NIST Handprinted Forms and Characters Database, www.nist.gov/srd/nistsd19.htm, 2007.

[7] Mitchell, T.M. – "Version spaces – an approach to concept learning", Report No. STAN-CS-78-711,Computer Science Dept., Stanford University, 1978.

[8] Mitchell, T.M. – "An analysis of generalization as a search problem", Proceedings IJCAI,6, 1979.

[9] Mitchell, T.M. – "Generalization as search.Artificial Intelligence", 18(2):203-226, 1982

[10] Quinlan, J.R. – "Induction of decision trees.Machne Learning", 1(1):81-106, 1982

[11] Shannon, C. – "A mathematical theory of communication.", Bell System Technical Journal,27:379-423, 1948

[12] Winston, P. H. – "Learning structural descriptions from examples", In P.H. Winston editor, 1975

[13] Winston, P. H. – "The psychology of Computer Vision ", New York, McGraw-Hill, 1975

[14] Winston, P. H. – "Artificial Intelligence", 3rd edition Reading, MA:Addison Wesley, 1992

[15] Sejnowski, T. J. and Rosenberg, C. R.– "Parallel networks that learn to pronounce English text.", Complex Systems, 1:145-168, 1987

[17] Hecht-Nielsen, R. – "Counterpropagation networks", Applied Optics, 26:4979-4984, 1984

[18] Qun Z., Principe J.C. – "Incorporating virtual negative examples to improve SAR ATR", Proceedings of the SPIE – The International Society for Optical Engineering, v 4053, 2000, 354-360

[19] Principe J.C., Dongxin X., Qun Z. – "Learning from examples with information theoretic criteria",Journal of VLSI Signal Processing System for Signal, Image and Video Technology, v 26, 2000, 61-77

[20] Qun Z., Principe J.C. – "Improve ATR performance by incorporating virtual negative examples", Proceedings of the International Joint Conference on Neural Networks, 1999, 3198-3203

Synthesis of Optimal Control Systems: A Comparison Between Model Checking and Dynamic Programming Techniques

Giuseppe Della Penna, Daniele Magazzeni
Department of Computer Science
University of L'Aquila
L'Aquila, Italy
Email: {dellapenna,magazzeni}@di.univaq.it

Benedetto Intrigila
Department of Mathematics
University of Roma "Tor Vergata"
Roma, Italy
Email: intrigil@mat.uniroma2.it

***Abstract*—In the last years, the use of sophisticated controllers, i.e., components that control larger systems, has become very common. In previous papers, we introduced a general, automatic methodology for the synthesis of optimal controllers which exploits *explicit model checking* to generate controllers for very complex systems. However, to completely validate the usefulness of our methodology we still need to compare it with other, well-established controller generation techniques that are currently applied to a variety *less complex* control problems. To this aim, in this paper we present a comparative evaluation, from both a theoretical and a computational point of view, between our model checking methodology and the widely applied dynamic programming approach, using the well known *inverted pendulum* problem as a test case. As a result, we find that our approach is *always feasible* when dynamic programming can be applied, and that the generated controllers *perform better* than the ones obtained by dynamic programming.**

I. INTRODUCTION

Control systems (or, shortly, *controllers*) are hardware/software components that control the behavior of larger systems, the *plants*. A controller continuously analyzes the plant state (looking at its *state variables*) and possibly adjusts some of its parameters (called *control variables*) to keep the system in a condition called *setpoint*, which usually represents the *normal* or *correct* behavior of the system.

In the last years, the use of sophisticated controllers has become very common in robotics, critical systems and, in general, in the hardware/software *embedded systems* contained in a growing number of everyday products and appliances. In particular, much work is being done to provide methodologies for the automatic (or semi-automatic) synthesis of controllers directly from the plant specifications.

In previous papers [1][2], we have introduced an automatic methodology for the synthesis of optimal controllers which exploits *explicit model checking* in an innovative way w.r.t. the approaches adopted so far. Moreover, looking for an *optimal* solution, our procedure actually can also be considered as a *planner*: indeed, it does not simply individuate good *local* actions, but searches for *the best possible sequence of actions* to drive the plant to the setpoint. Our methodology has been implemented in the CGMurφ tool [3].

The proposed methodology is general, and can be applied to nonlinear systems (see the case study in Section V) and hybrid systems (i.e., systems whose state description involves continuous and discrete variables). Indeed, in [1] we validated the effectiveness of our approach on the well-known *"truck and trailer with obstacles avoidance"* problem. This is a non trivial problem due to the (hybrid) dynamics of the truck-trailer pair and to the presence of obstacles which have to be avoided by the truck while maneuvering to reach the parking place.

In this setting, finding an *optimal* maneuver is a *very hard problem*, that cannot be modeled using common mathematical or programming strategies. This shows the potentialities of our methodology, but makes difficult a comparison with other, well established controller generation and planning techniques that are available and perform well on many kinds of *less complex problems*. In particular, a very natural question left open in [1] was a comparison with *dynamic programming techniques*, which are actually very suitable for the generation of optimal controllers [4],[5].

Therefore, to make a well defined comparison, we have chosen a case study which has been extensively worked out from a dynamic programming point of view: the *inverted pendulum* problem [6],[7]. Note that this apparently simple case study is instead an important issue in the controller design for many real-world systems. Indeed, many control problems, e.g., in engineering (i.e., the regulation of a steering antenna [8]) or robotics ([9]) can be reduced to an inverted pendulum problem.

In particular, in this paper we use [6] as a reference point, since it contains a complete theoretical treatment of the problem and also illustrates a detailed algorithm for the numerical synthesis of the corresponding controller. In this way, we are able to make a comparative evaluation, from both a theoretical and a computational point of view, between our methodology and the dynamic programming approach.

As a result, we find that our approach is *always feasible*, under the theoretical hypotheses established in [6] for the feasibility of the dynamic programming approach (see Section IV), and generates a controller that, in the case study under

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 54–59.

consideration, *performs better* than the one obtained through the dynamic programming approach (see Section V).

The paper is organized as follows. In Section II we describe our model checking based methodology and we give a short description of our tool, while in Section III we describe the dynamic programming based methodology. In Section IV we illustrate a theoretical comparison between two approaches and in Section V we present results of experimental comparison. Section VI concludes the paper.

II. Synthesis of Optimal Controllers by Model Checking Techniques

In this section we give a description of our model checking based controller generation technique and its implementation in the CGMurφ tool. Our technique applies very well to *hybrid systems* [2], but for the case study under consideration it is sufficient to specify it w.r.t. *continuous systems*.

Therefore, we suppose to have a plant P with continuous (discrete time) dynamics $F(x,u)$, where $x \in \mathbf{R^n}$ is the vector of the components of the state and $u \in \mathbf{R^m}$ is the vector of the components of the control.

The problem of the controllability of P to a specified state x_0, called the *setpoint*, is considered in a given bounded region H of $\mathbf{R^n}$, containing a neighborhood of x_0. Moreover, we assume that only a bounded set $U = \{u \in \mathbf{R^m}: ||u|| \leq \gamma\}$ of controls need to be considered.

After a suitable discretization of the continuous components of both the state and the control, we can assume that we have only a *finite* number of states and, in every state, only a *finite* number of allowed control actions. Note that the precision of discretization is a *design parameter* of our technique that the user can manage to get the controller best suited w.r.t. the general constraints she/he has to fulfill. Given a state x (a control u), we denote by $D_s(x)$ (resp. $D_c(u)$) the discretized value of x (resp. of u).

Our objective is to build an optimal controller for P, i.e., a controller that is able to drive P in the setpoint starting from any initial state. The optimality constraint requires the controller to reach the setpoint within the smallest possible number of steps (*time optimality*).

Thus, our controller has to decide, for every considered plant state, which is the best action (w.r.t. the number of steps) to reach the setpoint. The optimality of the action chosen implies the optimality of the generated controller.

In order to build such a controller, we consider the *transition graph* $\mathcal{G}$ of P, where the nodes are the reachable states and a transition between two nodes models an allowed action between the corresponding states. More precisely, $\mathcal{G}$ has a given set $\{s_0, s_1, \dots, s_l\}$ of *initial states* (that is, the states arising from the discretization of the region H). Moreover, $\mathcal{G}$ has a set of *goal states* G (that is, the (discretized) states sufficiently near to the setpoint). Finally, there is a transition u from the state s_i to the state s_j if u is a discretized control and the discretized value of $F(s_i,u)$ is s_j (i.e., $D_s(F(s_i,u)) = s_j$).

In this setting, the problem of designing the optimal controller reduces to finding the minimum path in $\mathcal{G}$ between each reachable state from initial ones and the nearest goal state. However, to limit the number of states to explore in this process, we impose some restrictions on the length of the path (time horizon). Such restrictions are realistic, since suitable response time limitations are always present in real-world control systems. On the role of time horizon limitation, see the next Subsection.

Unfortunately, the transition graph for complex, real-world systems could be often huge, due to the well-known *state explosion* problem. Thus it is likely that $\mathcal{G}$ would not fit into the available RAM memory, and then the minimum path finding process could be highly time consuming. However, model checking techniques [10] developed in the last decades have shown to be able to deal with huge state spaces. Thus, we used model checking algorithms, reshaping them in order to automatically synthesize optimal controllers, and we implemented this methodology in the CGMurφ tool [3].

In the following we describe our algorithm and we give a short description of the tool.

A. Algorithm for the Synthesis of Optimal Controllers

The algorithm for the optimal controllers synthesis takes as input the transition graph $\mathcal{G}$ and a time horizon T and outputs the controller table `CTRL`, containing all the states controllable to the goal in a suitable number of steps depending on T. Indeed, as a preliminary step, our algorithm performs a *reachability analysis* [11], an important step in the model checking methodology, starting from the initial states and iteratively performing all the possible control actions. In this process, we can very well leave the controllability region; it is therefore important to have a bound to the length of the sequences of states generated in the reachability analysis. In some fortunate case, where the state space in not too big, such a bound turns out not to be necessary: for every sequence we arrive at a point where all the possible actions bring back to already visited states. But in very complex cases ([12]) such time horizon restriction is unavoidable.

Our algorithm uses the following data structures:

- a Hash Table `HT` storing already visited states;
- a Queue `Q` containing the states to be expanded;
- a Predecessor Table `PT` containing, for each visited state, a list of its directed predecessors (note that we only need to store the position of a predecessor in the HT instead of the whole description);
- a Queue `Q_G` where we put the goal states encountered during the visit;
- the Controller Table `CTRL` containing (state p, action u) pairs. When the controller reads a state from the plant, it looks up the action described in the associated table entry and sends it to the plant.

The algorithm is organized in two phases: the first one (the `EXPLORE` procedure in Fig. 1) implements the reachability analysis performing a BF visit of $\mathcal{G}$ starting from all the initial states. Namely, the initial states are inserted in the queue `Q`. As usual in the BF strategy, we expand the state p in the front of the queue by computing the set of the successor states

```
1  EXPLORE(transition graph G, int T) {
2    foreach initial state p {
3      Enqueue(Q,p);
4      Insert(HT,p);
5      if (p ∈ G) Enqueue(Q_G,pos_into_HT(p));
6    }
7    while ((Q≠ ∅) && (current_BF_level ≤ T)) {
8      p := Dequeue(Q);
9      foreach s ∈ next(p) {
10       if (s ∉ HT) {
11         Insert(HT,s);
12         if (s ∈ G) Enqueue(Q_G,pos_into_HT(s));
13         else Enqueue(Q,s);
14       }
15       addPredecessor(PT,pos_into_HT(s),pos_into_HT(p));
16 } } }
```

Fig. 1. The EXPLORE procedure

$next(p) = \{D_s(F(p,u)) | u \in U \text{ and } u = D_c(u)\}$ (where the condition $u = D_c(u)$ implies that we are considering only discretized controls). For each successor s, if s is a new state, we mark it as visited by putting it in HT. Moreover, if s is a goal state we insert it in the queue of the goals Q_G, else we insert s in the queue Q of states to be expanded in the next level of the BF. Finally, we store the state p as direct predecessor of all states in $next(p)$ by adding its HT position in the predecessors list of such states. This phase ends when the queue Q is empty or when the maximum level of the search is reached and it outputs the predecessor table PT.

In the second phase, the SYNTHESIZE procedure (Fig. 2) visits the predecessor table PT in breadth first order starting from the goals and builds the controller table CTRL (initially empty). Namely, the BF visit queue Q is initialized with the goal states in Q_G. When a state s is extracted from Q, we find the corresponding entry in PT and we consider all its predecessors $p_1, \cdots, p_k$. For each pair (p_i, s) we find the discretized control action $u \in U$ such that $D_s(F(p_i, u)) = s$ and we assign the action u to p_i only if p_i is not present in CTRL (note that the use of the BF strategy in both phases, guarantees the time optimality of u). This phase ends when the queue Q is empty or when the maximum level of the search is reached, resulting in the final controller CTRL.

```
1  SYNTHESIZE(transition graph G, int T) {
2    Q := Q_G; //this erases the previous content of Q
3    while ((Q≠ ∅)&& (current_BF_level ≤ T)) {
4      pos_s := Dequeue(Q);
5      state_to := HT[pos_s].state;
6      foreach pos_pred in PT[pos_s] {
7        state_from := HT[pos_pred].state;
8        U := {u ∈ U | D_s(F(state_from, u_c)) = state_to};
9        local_action := pick an action in U;
10       if (CTRL[pos_pred] is not defined) {
11         addTransition(CTRL,state_from,local_action);
12         Enqueue(Q,pos_pred);
13 } } } }
```

Figure 2: The SYNTHESIZE procedure

B. The CGMurφ Tool

We implemented our technique in the Controller Generator Murφ tool [3], that is an extended version of the CMurφ [13] explicit model checker where the standard verification algorithm is replaced with the implementation of our technique.

CGMurφ describes the plant behavior by means of a collection of guarded transition rules, whereas the *goal construct* is used to define the set of goal states, i.e., the states that the controller should to bring (or maintain) the plant to.

Moreover, in order to generate controllers for complex and hybrid systems, CGMurφ allows to use finite precision real numbers (to model systems with continuous variables) and external C/C++ functions (to easily model the most complex aspects of the plant, or even interface the model with a complete plant simulator).

The output of CGMurφ is a *controller table* that contains, for each reachable system state s that leads (in one or more steps) to a goal, a pair (r,k) indicating that the shortest path leading from s to the nearest goal state has k steps, where the first step is the action given by rule r.

Finally, the tool provides a general validation test that allows to check the effectiveness of the discretization used for the synthesis, and eventually to repeat it with a finer one.

III. Synthesis of Optimal Controllers by Dynamic Programming Techniques

In the following of the paper we assume acquaintance with [6]; however, for the sake of clarity, in the present Section we briefly recall the main characteristics of the approach.

Let as before P be a plant with continuous dynamics $F(x,u)$, where $x \in \mathbf{R^n}$ is the state and $u \in \mathbf{R^m}$ is the control. The problem of the controllability of P to the *origin* (i.e., the setpoint) is considered in a given bounded region G, containing a neighborhood of the origin itself.

The classical dynamic programming approach proceeds as follows. First an optimal cost function J is considered:

$$J(x) = \inf_{\underline{u}}[\textstyle\sum_{t=0}^{\infty} l(F(x,u_t),u_t)] \tag{1}$$

where $l(x,u)$ is a continuous, positive definite *cost function* and $\underline{u}$ stands for a generic control sequence: $\underline{u} = \{u_0, u_1, u_2, \ldots\}$.

J is well defined (i.e., the infimum always exists in the region of interest) if and only if the plant P is controllable. In this case, J satisfies the so-called *Bellman Equation*:

$$J(x) = \inf_{u}[l(x,u) + J(F(x,u))] \tag{2}$$

and it can be computed by the following iterative method:

$$\begin{aligned} &J_0 = 0 \\ &J_{T+1}(x) = \inf_{u}[l(x,u) + J_T(F(x,u))] \end{aligned} \tag{3}$$

where $T \in \mathbf{Z}_0^+$.

Since the convergence of (3) poses several problems, in [6] the numerical design of the controller is based on a set of simplifying assumptions. Indeed, in [6] the authors assume to have a continuous, positive definite *terminal cost function*

$\bar{V}(x)$ and that there exists a bounded region $\bar{\Gamma}$, which includes G, such that:

$$\forall x \notin \bar{\Gamma}, u \in \mathbf{R^m}.\ l(x,u) > \bar{V}(x) \quad (4)$$

Other conditions are imposed on both $l(x,u)$ and $\bar{V}(x)$, that here are omitted for brevity.

With these assumptions, it is possible to define an *extended cost function* $V(x)$ as follows:

$$V(x) = \inf_{\underline{u}, t' \in \mathbf{Z_0^+}} [\sum_{t=0}^{t'-1} l(F(x,u_t),u_t) + \bar{V}(F(x,u_{t'}))] \quad (5)$$

where $\underline{u}$ stands for a generic control sequence: $\underline{u} = \{u_0, u_1, u_2, \dots\}$.

Thus, $V(x)$ goes trough all possible *finite time horizons* with terminal cost $\bar{V}$ and takes the infimum. The role of functions $l(x,u)$ and $\bar{V}(x)$ in (5) can be informally explained as follows: by the assumption (4) above, the final cost $\bar{V}$ can be considered as a *penalty function* which *punishes* a wrong control move; it regulates the convergence of the iterative computation of $V(x)$, since such iteration either stops with some final cost or the final cost diminishes further and further, as the minimal cost sequence drives to (a small neighborhood of) the set point.

The main results in [6] are the following:

Proposition 3.1: *$V(x)$ is continuous and satisfies the following Bellman-like equation:*

$$V(x) = min\left\{\bar{V}(x), \inf_u [l(F(x,u_t),u) + V(F(x,u))]\right\} \quad (6)$$

Now let $\gamma = \sup_{\xi \in G} V(\xi)$ and $\Gamma' = \{x \in \mathbf{R^n} | V(x) \leq \gamma\}$. Γ' is the region where we expect that control trajectories (if they exist) evolve.

Proposition 3.2: *If for every $x \in \Gamma'$, $V(x) < \bar{V}(x)$ holds, with the exception of a (small) neighborhood $N(\rho)$ of the origin, then in G the plant P is controllable in the sense that it can be driven to the neighborhood $N(\rho)$ of the origin. Moreover the control sequence can be determined as the minimal cost trajectory, from the equation (6).*

IV. A Theoretical Comparison

The basic theoretical precondition for the applicability of our methodology is the knowledge of some suitable upper bound to the search procedure that implements the *reachability analysis*.

Indeed, starting with some initial state x_S, the search for a control sequence driving x_S to the set point, can go indefinitely on, as x_S can turn out to be not controllable. In this sense the existence of an upper bound, for every state in the region of interest, is *a necessary and sufficient condition for the synthesis of the controller* through our methodology. Moreover, even if we know that a given x_S is controllable, an upper bound is crucial to reduce the search space.

In this Section, we want to show that under the conditions established in [6] for the feasibility of the dynamic programming approach in the synthesis of a numerical controller, the mentioned upper bound can be evaluated, from these assumptions, for every state x belonging to the region of interest. So, in such conditions, also our approach is viable.

To show this, we borrow from [6] the following result (see Section III for notation and terminology). For the computation of $V(x)$ in $\bar{\Gamma}$, only a bounded sets of controls needs to be considered:

$$\bar{U} = \{u \in \mathbf{R^m} | \|u\| \leq \gamma_1\} \quad (7)$$

where γ_1 is a suitable constant (see [6] III.B.(iii)).

Now we can prove the following proposition:

Proposition 4.1: *Fixed an open neighborhood of the origin: $N(\rho) = \{x \in \boldsymbol{R^n} | \|x\| < \rho\}$, the cost function $l(x,u)$ attains a positive minimum in the set $(\Gamma' - N(\rho)) \times \bar{U}$.*

Proof. Indeed $l(x,u)$, being a continuous function, attains a minimum in the closed, bounded set $(\Gamma' - N(\rho)) \times \bar{U}$. Let (x_M, u_M) this point.

By condition III.B.(12) of [6], we have that: $l(x_M, u_M) \geq \underline{\varphi_l}(\|x_M\|)$, where $\underline{\varphi_l}: \mathbf{R_0^+} \to \mathbf{R_0^+}$ is a strictly increasing, continuous function, such that $\underline{\varphi_l}(0) = 0$. So $\underline{\varphi_l}(\|x_M\|) > \underline{\varphi_l}(\rho) > 0$.

Let δ be the positive minimum of $l(x,u)$. Now let any state x be given in the controllable region Γ'. An upper bound b_x to the length of a control sequence for x, driving x to the neighborhood $N(\rho)$ of the set point, is given by the following equation:

$$b_x = V(x)/\delta \quad (8)$$

So we have proved the following proposition, for the conditions stated in [6]:

Proposition 4.2: *If the controller synthesis is possible by the dynamic programming approach then it is also possible by the model checking approach.*

Of course Equation (8) gives just an upper bound. An heuristic value, for tuning the search space, is possibly given by the mean value of $l(x,u)$.

V. An Experimental Comparison

In order to provide an experimental comparison of the two methods, we considered the inverted pendulum problem according to the formulation presented in [6].

More precisely, we are interested in the number of steps required to bring the system from the initial positions to a goal. To this aim, we implemented the algorithm proposed in [6] and we synthesized a CGMurφ-based controller.

In the following, we first describe the inverted pendulum problem and then we present the experimental results and compare them.

A. The Inverted Pendulum Problem

A controller for the inverted pendulum (Fig. 3) has to bring the pendulum to equilibrium by applying a torque in the shaft.

We describe the pendulum state with two real variables:

- the pendulum angle w.r.t. the vertical axis θ
- the angular velocity $\dot{\theta}$

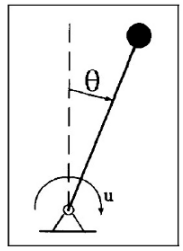


Fig. 3. Inverted Pendulum

The normalized system is described by the differential equation

$$\begin{aligned} \dot{x}_{(1)} &= x_{(2)} \\ \dot{x}_{(2)} &= \sin x_{(1)} + g(u) \end{aligned}$$

where $x_{(1)} = \theta$, $x_{(2)} = \dot{\theta}$, u is the torque in the shaft and

$$g(u) = \begin{cases} -0.7 & for\ u \le -0.7 \\ u & for -0.7 < u < 0.7 \\ 0.7 & for\ 0.7 \ge u \end{cases}$$

As observed in [6], being the torque bounded to a small amount, the pendulum cannot directly be turned from the hanging position into the upright position. Instead, it is first necessary to gain enough momentum. As a consequence, the dynamics of the system is *non-linear* and moreover, as proved in [6], the system *cannot be stabilized by a continuous controller*. So all the methodologies that produce only continuous controller would fail.

The model parameters are in Table I.

TABLE I
INVERTED PENDULUM PARAMETERS

θ_{min}	-6.6 rad	θ_{max}	6.6 rad
$\dot{\theta}_{min}$	-3.6 $\frac{rad}{sec}$	$\dot{\theta}_{max}$	3.6 $\frac{rad}{sec}$
u_{min}	-0.7	u_{max}	0.7
d_1	0.06	d_2	0.07
T_s	0.5	$\bar{T}$	50

In particular, d_1 and d_2 represent the discretizations for the state space and for the control space, respectively, T_s is the sampling time and $\bar{T}$ is the number of iterations, used only in the dynamic programming algorithm.

B. Inverted Pendulum Controller Synthesis by CGMurφ

We defined the CGMurφ model of the inverted pendulum according to parameters shown in Table I, thus we have the same conditions in both methods under comparison. Table II shows the results of the synthesis.

TABLE II
EXPERIMENTAL RESULTS FOR CONTROLLER SYNTHESIS

Reachable States	Rules Fired	Controller Transitions	Time (sec.)	Controller Size
26741	561371	20991	12	209 Kb

Note that the size of our controller is slightly bigger than the one in [6], due to larger information we save.

In order to check that the discretization we chose is sufficiently fine we performed a *validation test*. Namely, for each state s in the controller table we compared the pseudo-trajectory generated within the discretization d_1 with the real-trajectory generated using the full machine precision. So we say that s is controllable if the lengths of the two trajectories differ less than a given tolerance.

Moreover, to check the *robustness* of the controller, we repeated the validation test on a larger set of states. Namely, we randomly selected, for each state s in the controller table, 10 states in the neighborhood of s (i.e., states p such that $D_s(p) = s$). Table III shows the results of the validation test. Note that also with a small tolerance (5%), we obtain a completely satisfying result that states that our methodology is able to synthesize *effective* and *robust* control systems.

TABLE III
VALIDATION OF THE INVERTED PENDULUM CONTROL SYSTEM

Tol	Controlled States	Uncontrolled States	Random Controlled States	Random Uncontrolled States
5%	7278 (98%)	147	72769 (98%)	1481
10%	7366 (99.19%)	59	73607 (99.13%)	643
20%	7408 (99.77%)	17	73943 (99.59%)	307
30%	7422 (99.95%)	3	74022 (99.69%)	228
35%	7423 (99.96%)	2	74056 (99.74%)	194
47%	7425 (100%)	0	74203 (99.94%)	47
54%	7425 (100%)	0	74250 (100%)	0

C. Inverted Pendulum Controller Synthesis by Dynamic Programming

Here we recall the design algorithm derived from theoretical results presented in [6] and summarized in Section III and then we show its application to the inverted pendulum problem.

Given the discrete set of points $z_1, z_2, \dots, z_{\bar{n}}$ contained in $\bar{\Gamma}_D$ and the discrete set of control actions $u_1, u_2, \dots, u_{\bar{m}}$ contained in $\bar{U}_D$, we have the design algorithm shown in Fig.4.

```
Algorithm() {
 //initialization
 forall i ∈ [1, n̄]
  W_0(z_i) := V̄(z_i);

 for t=1 to T̄ {
  //Iteration from T to T+1
  forall i ∈ [1, n̄] {
   W_{T+1}(z_i) := min_{j∈[1,m̄]} {W_T(z_i), min[l(z_i, u_j) + I(W_T, F(z_i, u_j))]}

   t:=t+1;
}}}
```

Fig. 4. Design Algorithm

To apply this algorithm to the inverted pendulum, the design functions are chosen as follows:

$$\begin{aligned} \bar{V}(x) &= \min[80{,}640\,||x||^2] \\ l(x,u) &= \sqrt[4]{|x_{(1)}|} + \sqrt[4]{|x_{(2)}|} + 27|u|^3 + S_1 + S_2 \end{aligned}$$

where

$$S_1 = 80\left[\max\left(0, (|x_{(1)}| - 5.6)\right)\right]^2$$
$$S_2 = 80\left[\max\left(0, (|x_{(2)}| - 2.6)\right)\right]^2$$

D. Comparison of Experimental Results

We are interested in the number of steps required to bring the pendulum from a given initial position to the upright equilibrium position.

We simulated both controllers from the same initial positions and some of the results obtained are shown in Table IV. In particular, we present cases where $\dot{\theta} = 0$ that are the most complex ones, whereas in row (d) we have a simpler trajectory due to the fact that the required momentum is more quickly reached.

TABLE IV
COMPARISON OF EXPERIMENTAL RESULTS

	θ in rad	$\dot{\theta}$ in $\frac{rad}{sec}$	Steps Dyn.Prog.	Steps CGMurφ
a)	0.629	0.0	9	7
b)	0.7	0.0	21	19
c)	1.5	0.0	18	14
d)	1.5	-0.75	9	6
e)	3.50	0.0	10	8
f)	4.00	0.0	10	8

Table IV shows that the controller synthesized by CGMurφ always performs better. This is due to the fact that our methodology allows the analysis of a *much larger* number of trajectories, and therefore allows the choice of *better* trajectories. The model checking techniques (i.e., algorithms and data structures) make possible the computational treatment of the corresponding *huge* amount of data.

Indeed, by analyzing the corresponding pendulum trajectories, we observe that with the CGMur φ-generated controller there are less oscillations both around the upright position (to reach the equilibrium point) and the hanging position (to gain enough momentum). As an example, see Fig. 5 to be compared with Fig. 6b of [6].

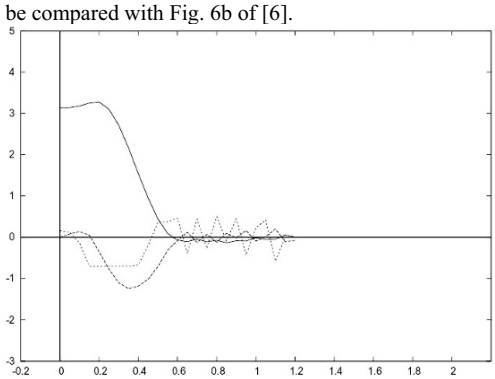


Fig. 5. Trajectory with initial condition $x = [\pi, 0]$ (x_1(——),x_2(− − −) and u(− · −) versus time).

VI. CONCLUSIONS

In this paper we have presented a comparison between our model-checking based methodology [1] for the synthesis of optimal controllers and a dynamic programming based methodology [6]. We have found that our approach is always feasible under the same conditions of [6] and, on a classical and representative case study, it performs better from a numerical point of view.

This last point can be explained observing that, by making use of the model checking techniques, we are able to explore a very large number of trajectories and therefore find better solutions. Moreover, our approach works directly on the plant dynamics, i.e., *without the help of auxiliary design functions*. This is an important point to be stressed: indeed, the mentioned design functions have to be find out case by case in the dynamic programming approach, whilst our approach produces a *press-button technology* that does not need any manual computation.

As a further research, we plan to make comparisons with other alternative approaches for non-linear systems, such as those coming from artificial intelligence, namely *neural networks* and *genetic algorithms*. Moreover, since our approach involves the analysis of a possibly huge number of states, we also want to consider the possibility of a *mixed approach* that uses other, lighter techniques, to drive the system reasonably near to the set point and then applies model checking to reach the set point itself.

REFERENCES

[1] G. Della Penna, B. Intrigila, D. Magazzeni, I. Melatti, A. Tofani and E. Tronci, "Automatic generation of optimal controllers through model checking techniques" in *Proceedings of 3rd International Conference on Informatics in Control, Automation and Robotics (ICINCO06), to be published in Informatics in Control, Automation and Robotics III, draft available at the url htt://www.di.univaq.it/magazzeni/cgmurphi.php*.

[2] G. Della Penna, B. Intrigila, D. Magazzeni, I. Melatti, A. Tofani and E. Tronci, "Automatic synthesis of robust numerical controllers", in *ICAS07*, IEEE Computer Society, p. 4.

[3] CGMurphi Web Page: *http://www.di.univaq.it/magazzeni/cgmurphi.php*.

[4] D.P. Bertsekas, *Dynamic Programming and Optimal Control*. Athena Scientific, 2005.

[5] M. Sniedovich, *Dynamic Programming*. Marcel Dekker, 1992.

[6] G. Kreisselmeier and T. Birkholzer, "Numerical nonlinear regulator design", *IEEE Transactions on Automatic Control*, vol. 39, no. 1, pp. 33-46, Jan. 1994.

[7] J. Hauser and H. Osinga, "On the geometry of optimal control: the inverted pendulum example", in *Proceedings of American Control Conference*, 2001.

[8] S. D. Ilcev, "Antenna systems for mobile satellite applications", *Global Mobile Satellite Communications*, pp. 175-234, 2005.

[9] K. Yokoi, F. Kanehiro, K. Kaneko, K. Fujiwara, S.Kajita and H. Hirukawa, "Experimental study of biped locomotion of humanoid robot HRP-1S", *Experimental Robotics VIII*, pp. 75-84, 2003.

[10] J.R. Burch, E.M. Clarke, L. McMillian, D.L. Dill and L.J. Hwang, "Symbolic model checking: 10^{20} states and beyond" in *Inf. Comput.*, vol. 98, no. 2, pp. 142-170, 1992.

[11] E.M. Clarke, O. Grumberg and D.A. Peled, *Model Checking*. The MIT Press, 1999.

[12] A. Ballucchi, L. Benvenuti, M.D. Di Benedetto, T. Villa and A. Sangiovanni-Vincentelli, "Idle Speed Control – a benchmark for hybrid system research", in *Proceedings of 2nd IFAC Conference on Analysis and Design of Hybrid Systems (ADHS06)*, 2006.

[13] CMurphi Web Page: *www.dsi.uniroma1.it/~tronci/~cached.murphi.html*.

[14] G. Della Penna, B. Intrigila, I. Melatti, E. Tronci and M. Venturini Zilli, "Exploiting transition locality in automatic verification of finite state concurrent systems", *STTT*, vol. 6, no. 4, pp. 320-341, 2004

An Artificial Immune System Based Multi-Agent Robotic Cooperation

Dioubate Mamady 1[1], Guanzheng TAN[1], Mohamed Lamine Toure[2], .Zeyad. M. Alfawaer[1]
[1]-*School of Information Science and Engineering, Central South University*
Changsha 410083, Hunan Province, P.R of China
[2]- *Department of Computer Science & Technology, Central South University*
Changsha 410083, Hunan Province, P.R of China
Email: dioubate24@yahoo.fr, tgz_csu@yahoo.com.cn, mohamedenchine@hotmail.com

Abstract:
This paper discusses three basic concepts: a) The behavioral management of artificial intelligence (AI) namely the intelligent multi agent systems, b) a geometric property of any object considered as an environment, and that define the average location of their weight and can completely describe the motion of any object "Uniform Mass, Common Shapes, Non-uniform Mass and a general shape" through space in terms of the *translation* of the center of gravity ($\mathcal{G}$) of the object from one place to another and the *rotation* of the object about its center of gravity if it is free to rotate. c) Artificial immune system that imitates the biological theory called the immune system and the evolutionary computation called Discrimination-based Artificial Lymphocyte (that includes a Genetic Artificial Immune System GAIS): modeling the learning mechanism of Self and Non-self Discrimination for environment idiosyncrasy. The outcome of this research is an Artificial Immune System based Intelligent Multi Agent robotic that solves agent-based applications. This new and specific strategy is applied to a robot cooperation problem focusing on the center of gravity where autonomous mobile robots emulate natural behaviors of our cells and molecules and realize their group behaviors; and the results prove that our method has solved the problem successfully.

Key words: Intelligent multi-gents, Artificial Lymphocyte categorization, Multi-Robot Cooperation.

1 - Introduction

Artificial immune system (AIS) imitates the natural immune system that has sophisticated methodologies and capabilities to build computational algorithms that solves engineering problems efficiently. One of the main objectives of AIS is to apply the immune metaphor or mechanism to computations [3]. In a natural immune system, recognition of the T cell is one of the most important mechanisms. It allows the immune system to discriminate self and non-self molecules. This characteristic is conducive to machine learning, particularly, classification and concept learning, in terms that one class (self) is distinguished from other classes (non-self).

In AIS study, however, the main topic has been the recognition mechanism and network theory of B-cells [21]. Although the mechanism of T cells is essential to immune discrimination, there are a few studies which this concept is applied to classification problems. Most of these studies either over simplifies fundamental processes [4] or are too complex by featuring GA with many parameters [8][9]. We need to research the generalized modeling of T cells and then the learning mechanism to apply these to machine learning.

Our work proposes a Multi-robot (agent) cooperation system utilizing the immune system, Artificial lymphocyte (*ALC)* and Genetic Algorithm (*GA)* concept where autonomous mobile robots emulate natural behaviors of ours cells and molecules and realize their group behaviors. Each robot contains sensors (like paratopes, idiotopes of antibodies improved by *ALC* as robot's possessor memory) to perceive a specification of antigen (Object) and other robots in several directions and decides its behavior based on the information obtained by the sensors.

With an aim of obtaining good linkages between natural and metaphor, we propose to exploit some characteristics of the natural systems related to the antigen "*antigen multi-determinate and multivalent" that presents several epitopes,* which means that the antigen can be recognized by several different B-cells. Hence, the same antigen can be recognized by several cells memories for the recognition of forms [7]. This natural fact increases the probability for a shape or a structure including $\mathcal{G}$ of being well classified. But, in general, determining the center of gravity ($\mathcal{G}$) is a complicated procedure because the mass (and weight) may not be uniformly distributed throughout the object. If the mass is uniformly distributed, the problem is greatly simplified. If the object has a line (or plane) of *symmetry*, the $\mathcal{G}$ lies on the line of symmetry. For a good multi robot cooperation that consist to move linearly a certain finite number of objects according their specific point like $\mathcal{G}$, we therefore attempt to use the mechanism of education in thymus, that define the functions of the adaptive immune system to recognize, eliminate and memorize foreign antigens, as well as to discriminate between self and non-self molecules. This discrimination phenomenon consists of *artificial lymphocytes*, a learning algorithm and a categorization method. Our method is applied to any object detection (its structure), motion problem; a specific application experimented to prove it. In this paper, an immune network is used for behavior decision controller and Dynamics that governs the system is given by equations in [16][15][25].

The motivation for this research comes from the fact that artificial immune system/artificial immune network has found solutions for several *robotic applications as well as optimization, neural network approaches, data mining* and *image classification applications* [14][16][13][11][24].

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 60–67.

2 - Natural Immune System

The natural immune system *(NIS)* is a complex network of cells, molecules, and organs. The *NIS* is capable of recognizing any foreign cell or molecule ("*antigens"*) and eliminating it from the body. To achieve this, it must perform pattern recognition tasks to distinguish molecules and cells of the body (called *"self"*) from foreign ones (called "*non-self"*). The basic components of the immune system are the white blood cells, called *self-cells* or *lymphocytes* in immunological terms. These specialized cells are classified into two types namely the *B lymphocytes (B cells)* and *T lymphocytes (T cells).*

- B *lymphocytes* are the cells produced by the bone marrows
- T cells develop in bone marrow and mature in *thymus.*

Definition1: When considered from a computational point of view, the immune system can be considered to be a rich source of inspiration or valuable metaphor as it displays learning, adaptability, is self-organizing, highly distributed, displays a memory and has no central point of control.

2.1 Properties of the Human Immune System.

The immune system possesses the properties from which the extensive survey of applications is presented as follows: The immune *response* [14]; *shape* of the epitope and complementary *match* phenomena [26][5]; Binding of a T-Cell's receptor to an MHC-molecules (*Major Histocompatibility Complex molecules*)[22]; *Positive and negative selection* [6]; *Clonal selection and expansion*[2][6]; *Immune memory*[5] and *Jerne's idiotopic network hypothesis*[17].

Cooperation between the different lymphocytes to detect an antigen occurs as explained in the section 3.3.3. Figure (1) shows the overall functioning of the immune system [23].

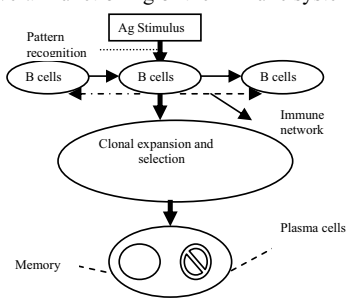


Fig 1: Representation of the human immune system.

*Definition*2: the basic agent definition as sketched out by, e.g., [12] states that an agent has the following characteristics:" it is a computer system that is situated in some environment and that is capable of autonomous action in this environment in order to meet its design objectives" where autonomy means control over behavior and internal state.

Multi agent systems (*MAS*) deal with the behavior management in collection of several independent entities, or *agents* [1]. Agents are entities with well-set goals, actions and knowledge in an environment that senses, communicates, coordinates, learns and makes decisions according to the environment.

Definition3: AIS are adaptive systems, inspired by theoretical immunology and observed immune functions, principles and models, which are applied to problem solving [4]

3. Description and aim of the problem:

We work with the paradigm idea and show how an agent may use concepts suggested by Artificial Immune Systems to dynamically change its intentions in response to a dynamically changing environment. A robot agent in a multi-agent scenario has to cooperate with other agents and is required to have capabilities to adapt and learn. The evolution of a robot agent is necessary for it to perform certain tasks even under unexpected environment changes. The goal of our method involves imitating the human immune system in terms of features and functions in multi robot system cooperation. The distinct similarities between the agent-robots and the immune system are: *Both are distributed or decentralized systems; Both have multiple autonomous entities; Both have individual and global goals; Both systems learn from their experience; Both are adaptable; Both sense the changes in the environment and act accordingly; Both systems communicate, cooperate and coordinate; Both possess knowledge with which they make intelligent decisions*. The immune system based autonomous multi robot architecture is derivable. The immune system is unique, robust, autonomous and multi-layered. It is augmented with a distributed leaning mechanism having lasting memory. It can have contextual recognition and noise tolerance. The behavior implementation of autonomous robots proposes that for a collective behavior to emerge, a set of agents must communicate with each other and with their environment.

3.1. Structure and properties of the antigens

In short, the antigens *Ag* are the molecules of the immune system. The epitope or antigenic determinant is the part of the antigen which interacts in a specific way with the antibody site. Dimensions of an epitope: *~ 0.7 x 1.2 x 3.5 nm* (the equivalent of approximately 5 to 7 *amino acids*) [20]. The same *Ag* molecule can be carrying several similar or different antigenic determinants. There are two antigens types: complete antigens and incomplete antigens.

3.1.1. Complete antigens: - can induce an immune reaction by themselves (e.g. a micro-organism).

3.1.2. Incomplete Antigens or haptens: A hapten is a small molecule which cannot stimulate by itself the *Ab* synthesis when this one is formed. These small molecules intervene not by their chemical nature but by their three-dimensional configuration.

- molecules of lower molecular mass which cannot induce immune reaction by themselves.
- When attached to a protein (proteinic conveyor), such as albumin or ferritin, a hapten can stimulate the production of specific antibodies. Once produced, the anti-hapten antibodies will recognize haptens even without the proteinic conveyor.
- haptens present generally only one epitope and they can exist spontaneously in nature (that is also the case of certain drugs).

The antigens are classified according to their characteristics of bonding (a total number of sites and a number of different sites):

3.1.3. Univalent and uni-determinate Antigen: It has only one epitope on the surface which is able to bind to an antibody.

3.1.4. Multivalent and unideterminate Antigen: It has at least two epitopes of the same type on an antigen molecule.

3.1.5. Multi-determinate and univalent Antigen: It presents several epitopes of various types, but only one of each type on an antigen molecule. The majority of the proteinic antigens fall into this category.

3.1.6. Multi-determinate and multivalent Antigen "MDMV Ag": Presents several and various types of epitopes; and more than an épitope of each type by antigen molecule. The proteins having multiple identical sub-units, as well as reticulate proteins and all cells fall into this category.

Attempts have been made to apply the Multi-determinate and multivalent Antigen idea to model a group of objects and to particularize the multi agent robot system cooperation.

3.2. Structure of Motion and Centre of gravity $\mathcal{G}$:

The *center of gravity* is a geometric property of any object. The center of gravity is the average location of the weight of an object. We can completely describe the motion of any object through space in terms of the *translation* of the center of gravity of the object from one place to another and the *rotation* of the object about its center of gravity if it is free to rotate.

In general, determining the center of gravity ($\mathcal{G}$) is a complicated procedure because the mass (and weight) may not be uniformly distributed throughout the object. The general case requires the use of calculus which we will discuss in the second part of this section. If the mass is uniformly distributed, the problem is greatly simplified. If the object has a line (or plane) of *symmetry*, the $\mathcal{G}$ lies on the line of symmetry. For a solid block of uniform material, the center of gravity is simply at the average location of the *M* physical dimensions.

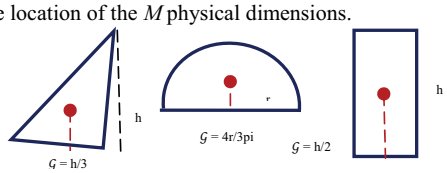


Fig 2: For Uniform Mass, Common Shapes

For a general shaped object, there is a simple mechanical way to determine the center of gravity: 1). If we just balance the object using a string or an edge, the point at which the object is balanced is the center of gravity. 2).Another, more complicated way. It's a procedure that works well for irregularly shaped objects that are hard to balance Ref **[10].**

If the mass of the object is not uniformly distributed, we must use calculus to determine center of gravity.

$$\mathcal{G} * W = \int x dw \qquad (1)$$

Where $\int \boldsymbol{xdw}$ is integration of a continuous function with respect to weight, x is the distance from a reference line, *dw* is an increment of weight, and *W* is the total weight of the object. To evaluate the right side, we have to determine how the weight varies geometrically. From the weight equation, the relationship between the mass *m*, volume *V* and density ρ of any object, we can solve the equation for the center of gravity if we have a functional form for the mass distribution:

$$\mathcal{G} * W = g \iiint x\rho dxdydz \qquad (2)$$

If we don't know the functional form of the mass distribution, we can numerically integrate the equation using a spreadsheet. Divide the distance into a number of small volume segments and determining the average value of the weight/volume (density times gravity) over that small segment. Taking the sum of the average value of the weight/volume times the distance times the volume segment divided by the weight will produce the center of gravity.

3.3. New adjustment mechanism

This paper seeks to apply CLONALG and ALC strategy to pattern recognition and assess its suitability for the domain. However, as the algorithm has never been applied to this task a before, a number of additional features have to be implemented.

3.3.1. To adapt our problem to the problem of structure of an antigen-object MDMV Ag" and its segmentation, we can divide the drive system of each class into several sub classes, gathering similar sub-classes (which have several joint characteristics) instead of using all the categories representing various sides of our object if, however, we could give it a regular geometrical shape. Thus, the training is carried out by using various examples of training per class (each class presenting a category of the epitope type). These categories represent a group of neighboring points to the center of gravity $\mathcal{G}$ and $\mathcal{G}$ itself. We wish to obtain for each class several representatives capable of capturing more characteristics of the class. It is after this procedure that the phase of appropriate side evolution using GAIS strategy starts. Therefore, we apply the algorithm proposed in [19] in which we introduce an algorithm of clustering to each type of classes to separate them in sub classes.

a) -The suggested solution:

Assume that each epitope or each group of epitopes of antigen-object is considered as if it was representing an antigen among an antigen population.

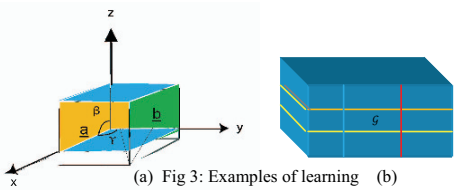
Step 1: Basic algorithm is: see [19] from which a set of antigenic patterns created represents "*MDMV Ag"*

Step 2: Clustering

In this step, we apply a clustering algorithm to the examples (of learning) of each epitope considered as a pattern of antigen.

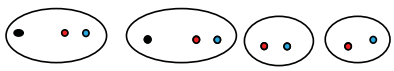
For each antigen-object:

Apply the clustering algorithm to the examples (of learning) of Ag (Figure 3a, 3b) to obtain a binary tree which represents the hierarchy of the found groups. To separate sub-classes, we apply a threshold TS to the obtained binary tree.



(a) Fig 3: Examples of learning (b)

This representation can take a simplified tree structure if the conditions of the motion are applied only to the two sides necessary for the displacement and to the sides parallel to the horizontal plane of the translation (negligible): four sides of possible impacts, two possible sides for the displacement containing the images of two (2) groups of points: a subgroup of points non neighboring to $\mathcal{G}$ and another subgroup of points neighboring to $\mathcal{G}$; this subgroup of points neighboring to $\mathcal{G}$ has also two subclasses: the center of gravity $\mathcal{G}$ and the nearest neighboring points "*NNP*" to $\mathcal{G}$; finally, two other sides whom is attributed the non possibility of direction of displacement, but containing the image of the centre of gravity . We obtain the following sub classes:



(c) Sub classes.

Step3: Learning

We use the same phase of learning used by CLONCLAS. Each sub class is treated like a class separate. As a result we obtain for each antigen Ag_i *(G+NNP)* a set of memories cells Ab_{mi}. Each memory cell belonging to Ab_{mi} has the same class as Ag_i.

Step 4: Classification

CLONALG generates the memory cells which will be used in the pattern recognition task. The cells are then presented with a new set of patterns to be classified.

b) - This stage of classification is only limited to classify the sides and *(G & NNP)* points (each side define a color and is represented internally as a binary string of length 120 bits) in sub groups. To privilege the direction of displacement linked cooperating agents' goal, we proceeded to the step 5; this step presents a new artificial immune system for classification (categorization) named discrimination-based artificial immune system (*DAIS*) containing Genetic Artificial Immune System GAIS and one the principle of self and non-self discrimination by T cells in the human immune system.

3.3.2. Objectives and Process of Cooperation.

Implementation and simulation process of our work need some principles of cooperation: We live in a world that is defined by three spatial dimensions and one time dimension. Objects move within this domain in two ways. An object *translates* or changes *location*, from one point to another. And an object *rotates*, or changes its *attitude*. In general, the motion of any object involves both translation and rotation. The translations are in direct response to external *forces*. The rotations are in direct response to external torques or moments (twisting forces). For *n* cooperating agents we have $\sum \vec{F} = M\overrightarrow{\gamma}$. Assume that environmental effects such as friction force $\vec{R}$, unknown disturbances, saturation, and so forth dependent on the system are negligible and the plan containing the translation is considered as horizontal. The above mentioned effects and speed phenomenon of the agent robots (phenomena being often explained by the *Jacobian matrix)* will be resolved in our future work.

3.3.3. Discriminate to categorize

To further specify the explanation of section 2: The important functions of the adaptive immune system are to recognize, eliminate and memorize foreign antigens, as well as to discriminate between self and non-self molecules. This means that the immune system is able to react only to non-self molecules. The adaptive immune system consists of two kinds of lymphocytes. One is the B-cell. B cells can recognize the shape of foreign antigens, eliminate and memorize them. The other is the T cell, which can discriminate between self and non-self molecules in a body. T cells acquire the ability to discriminate by education in an organ named the thymus.

What happens in the thymus? If the T cell recognizes self-molecules, it causes apoptosis and dies. This is called negative selection. If it does not react, it is tested with non-self molecules. If the T cell is not able to recognize them, it dies. So, T cells that can react with non-self molecules survive. This is positive selection. T cells that can pass through the thymus are ble to react to non-self molecules but unable to react to self molecules.

After training in a thymus, T cells circulate throughout the body. When B cells react with foreign antigens, they need the help of T cells. B cells do not react with self molecules if T cells do not recognize it. As a result, the adaptive immune system does not attach self-molecules.

Step 5: Self and non self discrimination for categorization.

We use a method used in [18] using a discrimination-based artificial immune system (DAIS), which consists of *artificial lymphocytes* (*ALC*), a learning algorithm and a classification method. But in our work the non-self class constitutes three sub classes, affinity threshold dependent of (α, β) and used side values represent string bits.

In the immune system, each T cell recognizes these molecules. AIS uses *artificial lymphocytes* (*ALC*) to recognize values individually in the same way. *The ALC consists of a receptor, an affinity threshold and a self class*. The receptor is represented in shape space. Its definition is such that, if there is a distance (Euclidian distance, hamming distance, etc.) between an ALC and individual value, the ALC can recognize it.

Assume that there are two kinds of region in shape space. Thymal education makes it possible for survived T cells to be distributed in only one region called the non-self region. The other region is named self region. Negative selection removes T cells from the self-region and positive selection chooses T cells distributed in the non-self region. This is very useful for categorization, because the values set belonging to the same class becomes new clusters for our case.

a - Basic learning strategy: This algorithm consists of three big parts: self and non-self classes, artificial lymphocyte and learning processes. Also, the value set is normalized. But in our case,

it is first necessary to divide the side set into two classes: self (self agent *SAc*) and non-self (non-self agent *NSAc*) classes. Self-class consists of only one class in the problem and non-self class consists of all the remaining classes (*NSAc* is divided in three subclasses: NSA_1; NSA_2 and NSA_3.But $NSA_3 \subset NSA_1+NSA_2$). The immune system categorizes self-molecules as self-class and all other molecules as non-self class. Side values belonging to self-class are named self data and those in the non-self class are named non-self data.

If class *i* is a self agent class (*SAc*) in a learning process, self class is *i*. Then, ALC_i represents the ALC who's self-class is *i* and the ALCs whose self class is the same are named the same kind of ALC.

Learning processes is based on education in a thymus and consists of iteration of some processes. These processes are *proliferation* and *mutation* (clonal expansion), determination of *affinity threshold AT* (D_{min} x α *(SAc)*; D_{min} x β *(NSA)* (negative selection) and *activation* and *selection* (positive selection). D_{min} represents minimal distance related to the recognition distance threshold of the artificial lymphocyte. After passing through these processes, the ALC_i is named *memory* ALC_i.

These processes make an ALC_i more effective. The processes are repeated to obtain the more useful ALC_i. It is the only one allowed to pass. This is repeated a predetermined number of times. The last selected ALC is stored as a memory ALC. *NAS* classes are evolved by GAIS strategy

Step 6: Genetic algorithm to assess the non-self agents' (NSA_1 and NSA_2) performance compared to SA.

- Interpret all kinds of genetic information, construct immune network, initialize environments (Uniform Mass and General Shape, cooperation's characteristics and conditions), and construct SA of agent ALCi.
- Applying the mechanisms of GA between NSA_1 and NSA_2. and Compare fitness values of the off springs to ALCi and exterminates the weakest agents.

After this process we'd get only one good of non-self class that will make affinity enough with *ALCi* of self class *i*. Therefore the two sub classes become only one; hence one activity. The algorithm thus becomes:

1. *Set class i to self class and j to non-self class.*
2. *Select one among the non-self data set, which the memory ALCis have not recognized. Remove the ALCs that cannot recognize the data selected (Activation).*
3. *Make an ALC_i whose receptor is the same as a selected non-self data.*
4. *Make ($Nc = 100$) clones of the ALC_i (Proliferation)*
5. *Mutate the receptor of these clones. Each element of the receptor is selected according to both predetermined mutation probability [0.2] and new characteristic δ of new non self population "NSA_3" from GA", and given a random value ([-0.1, 0.1]). (Mutation)*
6. *Select the ALC which has the biggest AT. Then, remove the rest. (Selection).*
7. *If the number of repetitions from Step 4 to Step 6 reaches Nc, go to Step 8. If not, return to Step 4.*
8. *Store the selected ALC in Step 6 as a memory ALCi.*
9. *If all non-self data are recognized by the set of the memory ALC_i, the learning is complete. Otherwise, return to point 2.*

We sow at first, class 1 becomes self class. Next, class 2 becomes self class. The process repeats this until number of classes *Ncl* becomes self class. This is the whole learning algorithm of new discrimination-based artificial immune system. After the learning process, there are all kinds of memory *ALCi*s that can distinguish between one class (their own self class) and remaining classes.

b- This stage of discrimination is limited to classify only the appropriate side and its ($\mathcal{G}$ & *NNP)* points in order to privilege the direction of displacement.

4. Application to Multi robot Cooperation

The problem implemented is Multi robot cooperation and the agent model describes the follow stages of processing namely *Pattern recognition, Binding process, Activation process, Post activation process and Post processing.* The experiment is simulated in MATLAB. The following section discusses the parameters used for this specific application.

4.1. Objects- Antigen and Antibodies and their characteristics:

A physical system is a set of material points which is considered as homogenously set in its mass or its volume. By analogy, we try an approach that considers the whole of the material points as the epitopes of an object- antigen (multi-determinate and multivalent antigen) of mass *M*, volume *V* and center of gravity $\mathcal{G}$. These epitopes are considered as uniformly distributed.

a) - The Object-antigen (*OAg*) is the objet to be move. Define the area to be explored for detecting the Object-antigen. This defines the boundary of the environment for the agents to locate the contact zone (good side). Objects are deployed in a uniform distribution within the environment. The initial locations correspond to the epitope that define the contact zone of the ($\mathcal{G}$&NNP). Characteristics of the *OAg are* stationary and have a regular geometrical shape. In this work we propose that there is not any interaction between epitopes or receptors of object-antigens. Choosing the good side of *OAg* is defined as the robot's good position for object motion (translation).

b) - Define the *B* cells (self-agent) to be the autonomous mobile robots. Deploy all the B-cells in a uniform distribution within the environment (which environment assume to be a matrix in which both the *B-Cell* and the *OAg* operate). The environment can be any one of the types of environment: *Deterministic* environment and *non- deterministic* environment; *Episodic* and *nonepisodic* environment depending on the application. We assume that there is an information vector for each *OAg*. We said above: When B-cells react with foreign antigens, they need the help of T-cells. B-cells do not react with self molecules if T-cells do not recognize it. Therefore, the initial locations of the B-cells correspond to the receptors (*ALCi*) of the agents. Characteristics of the B-cells are itinerant, independent, cooperative. The environment must be accessible and the robots get updated information about the environment.

Assume that all the robots have the capability to sense the objects and communicate between the agents within the idiotypic network (Euclidean distance measure for sides and D_m direction of displacement). Thus, every robot recognizes an

object and identifies its appropriate side location. On identification of the object every robot communicates to the other robot in a Jerne's network. For this problem, we have assumed Jerne's network as a broadcast network.

The agents are assumed to have sensory capability to identify the *OAg* within a region called *sensory neighborhood* [23]. They also possess the capability to communicate the *OAg* information to the other B-Cell within a region called *communication neighborhood*. In immune system, the communication circle is analogous to communication between B-cells connected in the immune network. In other words, every B-cell communicates the information to another B-cell that is within the communication neighborhood in the immune network as modeled in fig 1.

4.2. Simulation

The simulation describes emergent behaviors of autonomous mobile robots in dynamic and static systems. Any robot can identify the existence of other robots or object within a defined distance, and its behavior is to be front of *AOg* according the displacement sense. Each robot contains its sensory organ running as ALC, and its behavior uses this organ.

We assume that a priori knowledge of the object number with only one regular geometrical shape is known in the given environment. Therefore known number of objects is deployed in a uniformly distributed manner in the given area. We also deploy a known number of mobile robots in a uniformly distributed manner in the environment. The simulation differentiates the mobile robot and the object by using a red 'square' for an object and a blue 'o' for robots for representation while the code identifies an object by a '0' and the robot by a '1'. The information vector for the object and the robots contain the initially deployed location information along with the identifier using *ALC*. Table 1 shows an example of the object and the robot information vector. The simulation also requires setting the sensory circle of the robot and the communication circle. We have assumed that the communication circle is greater than the sensory circle.

	X coordinate	Y coordinate	Identifier using ALC
Object	5 4	6 8	0 0
Robot	3	4	1

Table 1: An Example of Information Vector of objects and Robots

The simulation is verified for the following variations.

- By increasing the sensory range from 3 to 9 units of distance measure.
- The communication circle was varied between 5 and 11 units of distance measure.
- Changing the environments area to 10 x10 and 32 x 32 rectangular grids. Here, the environment is accessible where each robot has the information about the both four sides and their *(G&NNP)* of *OAg* in the sensory and communication neighborhood. That is, on identification of the object, robot within the communication circle exchange about the sides of object detected and the appropriate side location for displacement, through the agent broadcast network. A sample environment vector is shown in Table 2. It can be seen from Table 2 that the robot 1 has the information about side 1 that is accessible to robot 2 if it is within the communication circle because robot 2 checks for the information available with robot 1 since it has not identified the good side. However, the environment becomes inaccessible on the assumption that the environment is not updated or when the communication circle is zero (com_cir = 0 and $D_m = 0$). It is useful to make the environment accessible in practice because, the mobile robots to be ready to move an object can be provided with the capability to communicate.
- . The memory was not used in this application problem since there is no usefulness in remembering the appropriate side (or the geometrical shape) of the object once it is located.

Index	Coordinates initial X Y	Identifier using com_cir	No of sides detected	Detected appropriate side using *ALC*
Objects	1 3 7 2 4 5	0 0	0 0	…. ….
Robots	1 2 4 2 5 2	1 1	4 0	1 ….

Table 2: An Example of the Environment Vector

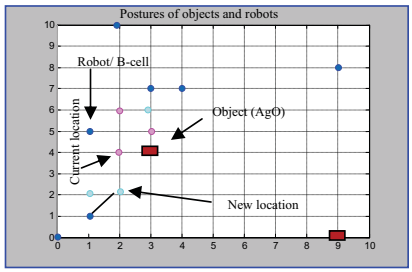


Figure 4: The locations of objects, function of sides, and robots after 4 iterations

The experiment is repeated for different populations of objects and robots. The typical range for the objects deployed are varied between 10 and 80 and accordingly and the robots are varied between 30 and 100. Figs: 4 and 5 show the simulation with objects and robots with their initial locations, $\mathcal{G}$ of objects and the three agents align themselves facing the appropriate side of the object.

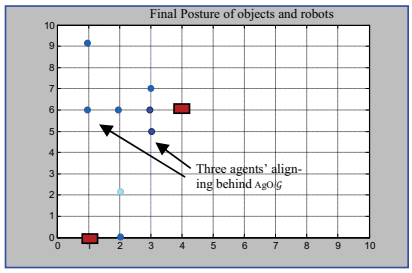


Figure 5: Three robots aligning themselves facing the appropriate side of the object after 12 iterations

4.3. Observations and evaluation

The following cases were studied and some results are discussed below.

a) For an increase in the sensory region and communication region the computational time in terms of rate of convergence was studied for both objects with uniform mass, common shapes and general shaped. Increasing the sensory and communication circles reduce the required the number of steps for the algorithm to converge. This is due to the fact that robots senses more area and can communicate with more robots and check if others have object information if they cannot find any.

b) For an increase in the population of objects (General Shape) and increase in population of robots the computational complexity in terms of rate of convergence (or the number of steps needed for the algorithm to less converge) was studied.

c) For an environment of same size and a constant sensory and communication circles, variation relationships of $\mathcal{G}$ for the convergence of both cases (uniform mass, common shapes and general shaped) and the number of iterations for a chosen value of robots and Objects are shown in Figures 6.

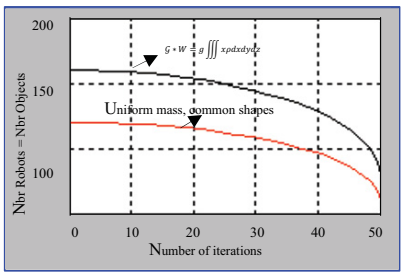


Fig: 6 variation relationships of $\mathcal{G}$ for both Uniforms Mass, Common shapes case and general shape case

5. Conclusion

This research is based on artificial immune system applicable to intelligent multi (agents) robots and provides to define the *Self-Angent "with ALC"* and *OAg* in terms of functions to be determined by the applications. Individual goals and a global goal for the agents can also be defined by the functions and an application is simulated. The object displacement problem using general shape is experimented and the results are better compare with uniform mass, common shape. This work merges the artificial immune method (artificial lymphocyte) in emergent behavior evolution of autonomous mobile robot as modeling the communication and cooperation system with the artificial immune network. In doing so in matrix environments, we construct an immune network representing robot and object and propose a model to evolve its structure. In this work, the main goal is to observe the communication and cooperation between robots for a common goal using immune network.

In addition, the motion of some objects is particularly complex because the rotations and translations are coupled together; a rotation affects the magnitude and direction of the forces which affect translations. For instance, in flight, both airplanes and rockets rotate about their centers of gravity. A kite, on the other hand, rotates about the bridle point. But the trim of a kite still depends on the location of the center of gravity relative to the bridle point, because for every object the weight always acts through the center of gravity. The motivation for this application is that in future, the displacement (linear or rotational…) or balance problem of the heavy object of which the $\mathcal{G}$ varies, can be performed efficiently by deploying mobile robots that have enough intelligence to detect and specify objects with high accuracy. To verify the generality of ours searches, more applications will be simulated including velocity and energy control idea quoted in (3.3.2). Thus, further research is needed to see if the proposed method in this paper can be applied into more various and complicated application areas in real world.

6. References

[1]. M. Wooldridge (1999). *Multiagent systems: a Modern Approach to Distributed Artificial Intelligence*, Gerhard Weiss (Eds.), The MIT press, Massachusetts, USA

[2].F. M. Burnet (1978). "*Clonal Selection and after*", In Theoretical Immunology, (Eds.), G. I. B. S. Perelson and G.H. Pimbley Jr., Marcel Dekker Inc, pp 63 – 85.

[3].De. Castro, L. and J. Timmis, 2001. Artificial Immune Systems: A New Computational Intelligence Approach. Springer

[4].De. Castro, L. and J. Timmis, 2002. Artificial immune systems: a novel paradigm for pattern recognition.

[5].L. N. Castro and F.J Von Zuben (1999). "Artificial Immune systems: Part I, Basic Theory and Applications", Technical Report – RT DCA 01/99, FEEC/UNICAMP, Brazil, 95 P.

[6].D. Dasgupta (1999).*Artificial Immune Systems and Their Applications*, Springer-Verlag, Germany.

[7].A. Deneche, S. Meshoul, M. Batouche Computer Vision Group, LIRE Laboratory, Université Mentouri University, Constantine, Algeria.

[8].A.J. Graaff, and A.P. Engelbrecht, 2003. Using a threshold function to determine the status of lymphocytes in the artificial immune system. Proc. SAICSIT, pp: 286-274.

[9].A.J. Graaff, and A.P. Engelbrecht, 2006. Optimized coverage of non-self with evolved lymphocytes in an artificial immune system. Intl. J. Computational Intelligence Res., 2: 127-150.

[10]. http://www.grc.nasa.gov/WWW/K-12/

[11].J. E. Hunt and A. Fellows (1996). "Introducing an Immune Response into a CBR system for Data Mining",In *BCS ESG'96 Conference* and published as Research and Development in Expert Systems XIII, 1996.

[12]. Wooldridge. Tractable Multiagent Planning for Epistemic Goals. *AAMAS'02*, July 15-19, 2002, Bologna, Italy

[13].G. W. Hoffmann (1986). "A Neural Network Model Based on the Analogy with the Immune System", j*ournal of Theoretical. Biology*, 122, pp. 33-67, 1986.

[14].P. Hajela and J. S Yoo (1999). "Immune Network Modeling in Design Optimization", In *New Ideas in Optimization*, (Eds.) D. Corne, M. Dorigo & F. Glover, pp. 203-215, McGraw Hill, London

[15].A. Ishiguro, Y. Watanabe, and T. Kondo (1997). "A Robot with a Decentralized Consensus-Making Mechanism Based on the Immune System", In Proc. ISADS'97, pp. 231-237.

[16].A. Ishiguro, Kondo, T, Watanabe, Y, Shirai, Y, and Uchikawa, Y, Emergent construction of artificial immune networks for autonomous mobile robots, in: IEEE International Conference on Systems, Man, and Cybernetics, Orlando, Florida, (1997), 1222-1228

[17].N. K. Jerne (1984). "Idiotypic networks and Other Preconceived Ideas", *Immunogocal review,* Vol.79, pp.5-24.

[18].Kazushi Igawa and Hirotada Ohashi: Discrimination-based Artificial Immune System: Modeling the Learning Mechanism of Self and Non-self Discrimination for Classification. Journal of Computer Science 3 (4): 204-

[19]. J. A. White and S. M. Garrett, "Improved pattern recognition with artificial clonal selection", In artificial immune systems, J.Timmis and al.(Eds): Proceedings of the 2nd International Conference on Artificial Immune Systems (ICARIS-03), LNCS 2787, pp. 181-193, 2003.

[20].McMurry & Castellion, Fundamentals of General, Organic & Biological Chemistry, 4*th ed,* 2003

[21]. J. M. Timmis, Neal and J. Hunt, 2000. An artificial immune system for data analysis. BioSystems, 55: 143-150

[22].A.S. Perelson, G. Weisbuch, "Immunology for physicists", *Reviews of Modern Physics*, vol. 69, no. 4, October 1997.

[23].Srividhya Sathyanath. Ferat Sahin. *AISIMAM* – An Artificial Immune System Based Intelligent Multi Agent Model and its Application to a Mine Detection Problem.2003

[24].S. Sathyanath and F. Sahin (2001). "Artificial immune Systems Approach to a Real Time Color Image Classification Problem", *Proceedings of the SMC2001***,** IEEE *international Conference on Systems, Man, and Cybernetics*, Vol 4, pp. 2285 – 2290, Arizona, USA.

[25]. Y. Watanabe, Ishiguro, A., Shirai, Y., Uchikawa, Y., (1998) "Emergent construction of behaviour arbitration mechanism based on the immune system", *Proc of ICEC 1998*, pp: 481-486. (1998)

[26].J. M. Timmis, Neal and J. Hunt (1999). "An Artificial Immune System for Data Analysis", *Proceedings of the International Workshop on Intelli gent Processing in Cells and Tissues (IPCAT)*, Indianapolis, U.S.A,

Towards ASSL Specification of Self-Scheduling Design and Monitoring in Team-Robotics Modeled with AS-TRM

Olga Ormandjieva and Emil Vassev
Department of Computer Science and Software Engineering
Concordia University
Montreal, Quebec, H3G 1M8, Canada
{ormandj, i_vassev}@cse.concordia.ca

Abstract- **Team-Robotics is an evolving mobile embedded system comprising a set of autonomous mobile robots operating in a dynamic environment and cooperating to fulfill a particular task. Scheduling is one of the main attributes of Team-Robotics. The objective of this research is to enable software self-scheduling and its monitoring at runtime in evolving Team Robotics systems modeled as reactive autonomic systems with the Autonomic Systems Timed Reactive Model (AS-TRM) architecture. In this paper, we propose a runtime self-monitoring methodology for controlling the correctness of behavior in such a system, based on the synchronous composition of its reactive autonomic components modeled as labeled transition systems, and from the specification of system policies such as safety and liveness. The AS-TRM architecture, sustaining the real-time self-scheduling modeling, verification and monitoring, is further refined with the Autonomic System Specification Language (ASSL) - a framework for formally specifying and generating autonomic systems. With ASSL, we can specify high-level behavior policies, thus it is a highly appropriate language for specifying self-scheduling mechanism as part of the overall team-robotics system behavior.**

I. INTRODUCTION

The number of mission-critical applications with a high cost of failure, or a high cost to fix, has increased exponentially in recent years, making the need for built-in behavioral monitoring greater than ever before. Autonomic computing is a research area led by the IBM Corporation, which focuses on building self-monitoring in complex computing systems [1]. This paper addresses the problem of self-scheduling modeling, verification and monitoring in Team-Robotics modeled with the Autonomic System Timed Reactive Model (AS-TRM) [2, 9], where each autonomic component is designed to react to every request from the environment in real time and synchronize its behavior with the environment (including other autonomic components).

Team-Robotics is an evolving complex mobile embedded system comprising a set of autonomous mobile robots operating in a dynamic environment to fulfill of a particular task. The robots are required to react in real-time to the requests for a task fulfillment or changes in their environment. The tasks must be scheduled, assigned to the mobile robots and optimized given robots' resource constraints, in real-time. The cooperation between the robots requires both communication and synchronization to ensure reliable performance of the supervised team of mobile reactive robots.

Modeling a schedule for team-robotics is a challenge for several disciplines such as software, electrical or mechanical engineering [6]. The objectives of this research are: 1) to develop an appropriate formal framework for continuous self-scheduling to support the distributed autonomic behavior of a team of synchronously interacting real-time reactive mobile robots, 2) to define an appropriate modeling approach for optimal scheduling of team tasks with given hard real-time timing constraints, system policies and resource constraints, and 3) to build corresponding architecture along with a communication mechanism to implement the self-scheduling of the distributed autonomic behavior of the team of robots.

One of the main contributions toward the achievement of the objectives stated above is the enabling of software self-scheduling and its monitoring at runtime through modeling Team-Robotics as reactive autonomic system with the Autonomic Systems Timed Reactive Model (AS-TRM) architecture. Another important contribution is the formalization of the self-scheduling mechanism in Team Robotics with ASSL [3]. The Autonomic System Specification Language (ASSL) is a framework that implies a multi-tier structure for specifying autonomic systems. By its virtue, ASSL is generic and expressive enough to describe a variety of autonomic systems [3, 4]. Moreover, ASSL focuses on the service-level objectives and self-management policies, thus making ASSL suitable for specifying self-scheduling and self-monitoring in Team Robotics. The advantages of using ASSL for modeling AS-TRM systems are that it enables software assurance and provides proof of correctness of the behavior of such systems. Moreover, formally derived models can be used as the basis for code generation.

The rest of this paper is organized as follows: Section II introduces the AS-TRM model for Team-Robotics. The methodology for self-scheduling is explained in section III and specified formally with ASSL in section IV. Section V introduces related work and justifies the proposed self-

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 68–76.

scheduling modeling approach. The conclusions and future work directions are outlined in ssection VI.

II. AS-TRM ARCHITECTURE OF TEAM-ROBOTICS

The section provides a comprehensive AS-TRM model of the Team-Robotics architecture, which is intended to capture and convey the significant decisions in the self-scheduling mechanism that will serve as a foundation for further schedule design, verification and self-monitoring of the scheduled tasks. In our approach, we assign the responsibility of scheduling and monitoring the team tasks to a Group Manager (GM) of the team of mobile robots. The autonomicity of the robots is preserved by their ability to provide self-control over their resources and the tasks' fulfillment. The behavioral models of the reactive autonomic robots RCRobot, their manager GMRobot, and the tasks (group GTP and local TP) are depicted visually in this paper in order to describe their behavior more clearly. The autonomicity of the robots is preserved by their self-awareness, as explained in the following section.

A. AS-TRM

The AS-TRM architecture for Team-Robotics (see Fig. 1) describes the Team-Robotics configuration as autonomic groups (teams) of reactive components (RC) Robots; the team behavior is managed by the Autonomic Group Manager (AGM) GMRobot. Autonomic functionalities such as self-scheduling and self-monitoring are implemented both i) at group level, by the group using the global policies such as scheduling and timing requirements, and ii) at reactive component, level using locally maintained knowledge on the robot's resources.

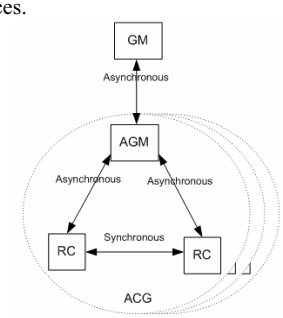
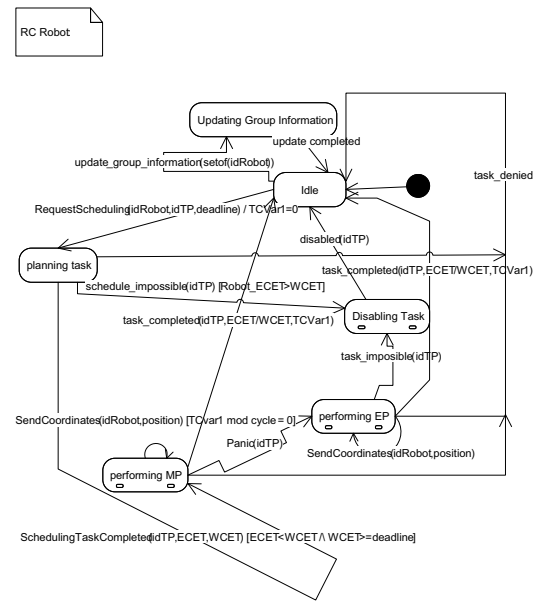


Figure 1. Sample AS-TRM Architecture for Team-Robotics

B. RC Robot

Mobile robots interact with a dynamically changing physical environment, which requires that they are self-aware and able to react appropriately while cooperating with each other to fulfill a task. At runtime, each robot reacts to the task assigned by the group manager (RequestScheduling event). The timing requirements of the task (deadline) imposed on the robot's behavior are expressed with the use of a local clock initialized when a new task is received by the robot. Each robot sends the team manager, GMRobot, the preprocessed output of its laser scanner to be fused with the data received from the other robots until object detection and localization has been achieved. Some team tasks would require broadcasting of the local results to the rest of the team members; for such cases, the required knowledge on the team members is available locally.

The self-awareness of a robot is expressed in its ability to self-control task execution: autonomicity requires highly effective local self-scheduling of the limited resources (namely CPU and communication bandwidth). Each robot plans for task execution based on its resources, and can disable a task if the timing requirements are not met. In such a case, the robot warns its supervisor (the team manager) by sending a *task_impossible* event. The robots' mobility requires a tracking mechanism for their supervision, which is implemented with a cyclic *SendCoordinates* task issued by the manager and received by the robots. The behavior of the reactive autonomic object modeling their self-awareness and ability to react appropriately is shown in Fig. 2.



g.2. Robot behavior

C. GMRobot

Cooperation between the robots in a team requires: (1) communication and synchronization, in order to avoid collisions, identify and localize objects [6], collect information (applying certain tools managed by the robots) and send the results to other robots on the team or to the manager; (2) scheduling of the sequence of tasks for communication between the robots; (3) definition and maintenance of common for the team timing and scheduling constraints. The above requirements call for a supervisor, who will be

responsible for managing and controlling the communication between the members of a team of mobile robots in such a way that all scheduling and timing constraints are met.

The team manager GMRobot performs the role of such a supervisor and has the following responsibilities: (1) auto-generation of a group schedule GTP to satisfy the requirements; (2) formal verification of the generated schedule against a set of consistent timing requirements and system policies; (3) reporting the results of a task execution to its superior (*TaskDone(GTP*) event). The scheduling is time-driven, and thus the self-scheduling must be performed safely in real time; a task can be prevented from occurring if timing requirements are not satisfied, which is modeled with the *task_denied* event issued by the manager, GMRobot, when the scheduling fails or when the GTP deadline is not met (see Fig. 3). In the case of a local task failure, the team manager delegates the task to an existing member of the group with similar characteristics, or when such a robot is unavailable, assigns a new robot (or robots) to the group (*rescheduling* state) to take over responsibility for fulfilling that task. The behavior of a team manager is modeled in the AS-TRM as a labeled transition system, as shown in Fig. 3.

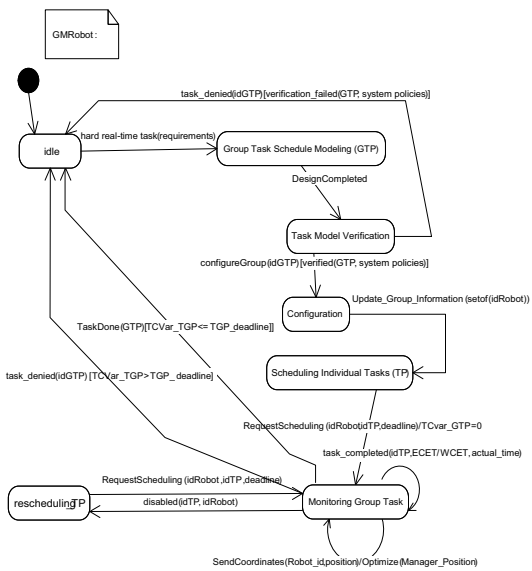


Fig. 3. GMRobot behavior

III. SELF-SCHEDULING

A team of mobile autonomous robots has to operate and cooperate in a dynamically changing environment under hard real-time constraints in order to fulfill a task. In this research, we focus on the specification of self-scheduling, which we consider to be a self-management policy. Self-scheduling in a team of autonomous mobile robots requires: (1) modeling of the schedule at group level (GTP) and at local level (TP); (2) monitoring of the scheduled tasks at both the individual level (RC Robot) and the group level (GMRobot). This paper addresses the problem of accomplishing an efficient and flexible formal model for scheduling tasks, which would allow for automatic verification of the tasks' synchronization and timing requirements, as well as their legal sequencing. The efficiency is achieved by the chosen underlying scheduling mechanism, namely, the Time-Aware Fault Tolerant (TAFT) scheduler [6]. Flexibility is built into scheduling self-monitoring by allowing task delegation between the collaborating robots and runtime reconfiguration of the team to meet the scheduling requirements. Enabled by the underlying communication layer, each task can take place on a different reactive component (RC Robot) for load distribution reasons or when a robot equipped with hardware adequate for this processing is unavailable [7]. The GMRobot is responsible for assigning and reassigning tasks to the members of the team. In this work, we make the assumption that there is an unlimited number of available robots for fulfilling the tasks.

A. Schedule Modeling

In our approach, a schedule is defined formally as a reactive autonomic object where the set of states represents the tasks to be scheduled, the set of events models the triggers for each task, the set of transitions models the required sequencing of tasks within the schedule and the set of timing requirements models the timing constraints on the scheduled tasks. The resulting sequence of states represents the task sequencing with the hard real-time requirements built into the model as guards on the transitions; the local clocks are employed for specifying the timing constraints.

Group Task (GTP) Schedule. Let a group-level new real-time task (GTP) consist of the following sequence of subtasks:

Task1. Collect data by laser scanners (robots)
Task2. Fuse the data and achieve a complete and accurate detection and localization of an object
Task3. Categorize the object

GTP contains a single, common deadline [6]. A local clock *TCVar_TGP* is initialized with the arrival of a new request for a hard real-time task, the deadline of which is predefined. The group task scheduling shown in Fig. 4 specifies the allowable sequences of tasks to be fulfilled by the team of robots under the supervision of the team manager.

Individual Task (TP) Schedule. GTP scheduling has to be further refined into TP scheduling of tasks assigned to individual robots, which are responsible for extracting the local data using their laser scanner equipment (see Fig. 5). The robots have to meet local deadlines for providing their output. Each individual robot is responsible for ensuring that tasks meet their deadlines by scheduling their own worst-case execution time (WCET) and expected case execution time (ECET).

B. Self-Scheduling Mechanism

The scheduling strategy we have adopted is the two-level Time-Aware Fault-Tolerant (TAFT) scheduler described in [6]. Moreover, we specify that the TAFT scheduling algorithm is part of the self-scheduling specification. The assumptions underlying the modeling of the tasks required to fulfill self-scheduling are:

1. The group leader (GMRobot) and the robots comprise all the execution routines needed to perform any task.
2. A request for performing a hard real-time task will provide the robots with a task ID of the task to be performed.

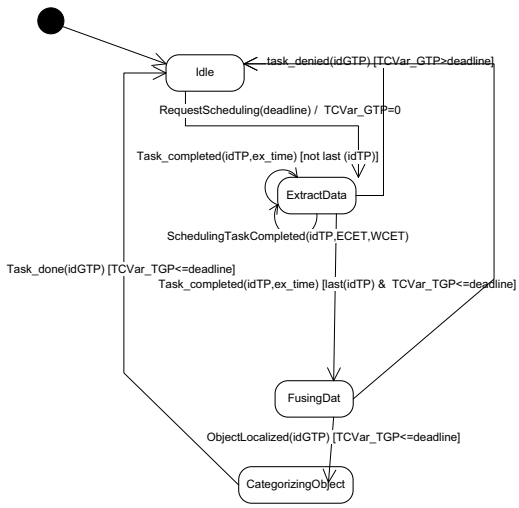


Fig.4. Group Task (GTP) specification

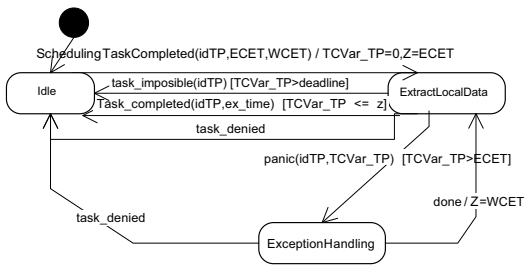


Fig.5. Robot Task (TP) specification

The following algorithm describes the self-scheduling policy specified with ASSL:

1) The *GMRobot* receives a group task-request message with a firm deadline from a superior robot.
2) The *GMRobot* models both group (*GTP*) and individual robot (*TP*) tasks and schedules them. If the scheduling is impossible or the time for modeling runs out the deadline, the *GMRobot* notifies its superior robot that the task is denied. Otherwise, *GMRobot* configures a team of RC Robots to fulfill the task.
3) The *GMRobot* sends a scheduling task request to all the robots, which should be guaranteed for a bounded time (deadline). The scheduling task has the form of independent task designed as a *TaskPair* (TP), which is a tuple with a *deadline*, a *MainPart* (*MP*), and an *ExceptionPart* (*EP*).
4) The robots compute their scheduling times. Each robot computes expected-case execution time (*ECET*) and worst-case execution time (*WCET*).
5) A scheduling will not be possible if a robot cannot compute its scheduling times, or if *WCET* < *ECET*.
6) If a robot cannot perform the scheduling task it reports "scheduling impossible" to the *GMRobot*
7) The rest of the robots perform a *TP* scheduling upon the scheduling task request from the *GMRobot*. Every robot should guarantee a maximum scheduling time equal to the deadline.
   a. A robot starts executing the *MP* of the scheduling task.
   b. If the execution time runs out the *ECET*:
      i. the robot switches to the *EP* of the scheduling task, thus bounded by the *WCET*;
      ii. if the execution time runs out the *WCET* the robot cannot perform the scheduling task and reports "task impossible" to the *GMRobot*;
      iii. in case of successful task completion, the robot notifies the *GMRobot* for "task completed within *WCET*" and for the actual task time.
   c. In case of successful completion of the *MP* the robot notifies the *GMRobot* for "task completed within *ECET*" and for the actual task time.
   d. The robot sends its position to the *GMRobot*.
8) If there are some robots not able to perform the scheduling task, i.e. they have sent a "task impossible" message, the *GMRobot* asks idle robots from the same group to perform the scheduling task. In case there are no idle robots from the same group, the *GMRobot* allocates new robots from other groups to perform the scheduling task.
9) The *GMRobot* finalizes the group scheduling task, by collecting all the "task completed" messages. The group scheduling task is successfully completed if all the "task completed" messages arrived within the group task deadline.

C. Schedule Verification

Verification of the schedule design is required to ensure that it is feasibility with respect to its correctness and consistency with team behavior before the actual execution of the GTP task. Our formal verification method is based on the synchronous composition of the RCRobot, GMRobot, GTP

and TP and on the system policies, a sample of which is provided below:

System_Policies <name>

Safety: *<vector_at(TP,ExtractLocalData,ExceptionHandling) ➔ object_at(GTP, ExtractData) > ...*

Liveness: *< vector_at(TP,Idle) /\ object_at(GTP, FusingData) ➔ occur(GTP,ObjectLocalized) \/ occur(GTP,task_denied)> ...*

End_System_Policies

The details of the synchronization generation algorithm as a means for achieving self-scheduling verification are omitted in this paper due to space constraints The synchronous behavior resulting from the algorithm can be mapped to all possible behavioral paths for the system. These paths can then be checked for consistency with the requirements for legal behavior of Team-Robotics, which will add value to the scheduling correctness self-control procedure.

The AS-TRM architecture, sustaining the real-time self-scheduling modeling, verification and monitoring, is further refined with the Autonomic System Specification Language (ASSL) - a framework for formally specifying and generating autonomic systems.

IV. ASSL Specification of Self-Scheduling

The ASSL framework is defined through formalization tiers. Over these tiers, ASSL provides a layered structure for specifying autonomic systems as formal executable models. ASSL defines an AS with its interaction protocol and autonomic elements (AE). The ASSL tiers and their sub-tiers describe different aspects of the AS, like policies, communication interfaces, execution semantics, actions, etc. The elements representing the major tiers and sub tiers in ASSL are described in [3, 4]. All of them ensure that the system is well-defined and consistent, providing a "bottom-up" style where the upper tiers are expressed using the elements from the lower ones.

A. ASSL Tiers

The ASSL specification model provides three main tiers, each containing a few sub tiers [3, 4], which describe different aspects of the AS, like policies, communication interfaces, execution semantics, actions, etc. The following elements briefly describe the ASSL tiers.

AS Tier. The AS tier specifies an AS in terms of service-level objectives (AS SLO), self-management policies, metrics and architecture. The AS SLO is a high-level form of behavioral specification that establishes system objectives such as performance. The self-management policies could be the four "self-" policies of an AS, as stated in [1]: self-configuring, self-healing, self-optimizing and self-protecting. Metrics are a set of parameters and observables controllable by the AEs.

AS Interaction Protocol Tier. At this tier level, the ASSL framework specifies an AS-level interaction protocol (ASIP). The ASIP is a public communication interface, expressed as messages exchanged among AEs using communication channels and communication functions.

AE Tier. The ASSL formal model considers AEs to be analogous to software agents able to manage their own behavior and their relationships with other AEs. Here, ASSL describes the individual AEs of the AS.

B. Self-Scheduling Specification with ASSL

As a formal language, ASSL defines a language-independent representation for Autonomic systems, where they are described as a set of interacting AEs. ASSL provides a rich set of structures and elements, including selfmanagement policy structures [3, 4].

AGM Manager of Group of Robots.

At this tier, we specify a global resource metric, called *numberOfIdleRobots* (See Listing 1), which, as its name implies, is an indicator of the number of idle robots in the entire system. This metric is used by the group leader's self-scheduling policy (see Section III) to check the availability of idle robots, when there is a lack of idle robots in the group. Moreover, at this tier, we specify the Team-Robotics ASTRM architecture model. In our example, we consider a system consisting of a *SuperiorRobot*, a *GMRobot* (the group leader) and three robots. As specified (see Listing 1), *RoboticsASTRM* has a centralized architecture. The *SuperiorRobot* is the one who requests group tasks to be performed. It is not specified in our example. The *GMRobot* and all the robots are grouped together, in *groupRobots*, where the *GMRobot* is the group council node [3, 4], and *Robot1*, *Robot2* and *Robot3* are the member nodes. The group council is an AE coordinating the work of the group.

```
AS RoboticsASTRM {
  METRICS {  METRIC numberOfIdleRobots {
        type = RESOURCE;
        description = "The number of idle robots in the entire AS.";
        THRESHOLD_CLASS: integer [ 0, ); } }
  ASARCHITECTURE { // centralized system with 3 robots
     AE_LIST: SuperiorRobot, GMRobot, Robot1, Robot2, Robot3;
     DIRECT_DEPENDENCIES { DEPENDENCY SuperiorRobot: GMRobot;
        DEPENDENCY GMRobot: Robot1, Robot2, Robot3;
        DEPENDENCY Robot1: GMRobot;
        DEPENDENCY Robot2: GMRobot;
        DEPENDENCY Robot3: GMRobot }
     TRANSITIVE_DEPENDENCIES {};
     GROUPS {    GROUP groupRobots {
           MEMBERS: Robot1, Robot2, Robot3;
           FINAL COUNCIL: GMRobot; } }  } // ASARCHITECTURE   }
```

Listing 1: AS Robotics ASTRM Partial Specification

ASIP Tier

At this tier, we specify the communication protocol needed by the *GMRobot* and the robots to communicate and transfer data for the needs of the self-scheduling policy. First, we specify the messages needed for the self-scheduling specification (Listing 2):

- *groupTask* - used by the *SuperiorRobot* to request from the *GMRobot* to perform a group task with a firm deadline. The message contains the task deadline and the task ID.
- *requestScheduling* – used by the *GMRobot* to request from the robots to perform a scheduling task. The message contains the task deadline and the task ID.
- *taskCompletedECET/taskCompletedWCET* – used by the robots to notify the *GMRobot* that they have completed their task within ECET/WECT.

- *taskImpossible* – used by the robots to notify the *GMRobot* that they cannot complete their task within the deadline time.
- *robotPosition* – used by the robots to notify the *GMRobot* for their current position.

In addition, at this tier, we specify a single communication channel (see *schedulingChannel* in Listing 2) and the functions operating the messages over that channel. It is important to mention that the ASIP specification is not complete. In Listing 2, we present only those messages, channels and functions needed by the system self-scheduling specification.

```
ASIP {
   MESSAGES { MESSAGE groupTask {
         SENDER: { SuperiorRobot };
         RECEIVER: { GMRobot};
         TYPE: TASK;
         PARAMS: {double deadline; integer idTP }
         BODY: {"groupTask"} }
      MESSAGE requestScheduling {
         SENDER: { GMRobot };
         RECEIVER: {Robot1, Robot2, Robot3};
         TYPE: TASK;
         PARAMS: {double deadline; integer idTP}
         BODY: {"schedulingTask"} }
      MESSAGE taskCompletedWCET {... }
      MESSAGE taskCompletedECET {....}
      MESSAGE taskImpossible {....}
      MESSAGE robotPosition {....} }
   CHANNELS {   CHANNEL schedulingChannel {
         ACCEPT: { requestScheduling, taskImpossible, robotPosition,
                   taskCompletedWCET, taskCompletedECET };
         ACCESS: SEQUENTIAL;
         DIRECTION: BIDIRECTIONAL } }
   FUNCTIONS {
      FUNCTION sendTaskCompletedWCETMsg {.... }
      FUNCTION sendTaskCompletedECETMsg {....}
      FUNCTION sendTaskImpossibleMsg {....}
      FUNCTION sendNewRobotAllocatedMsg {....}
      FUNCTION sendRobotPositionMsg {....}
      FUNCTION sendRequestSchedulingMsg {....}
      FUNCTION receivegroupTask {....}
      FUNCTION receiveTaskCompleted {....}
   } } // ASIP
```

Listing 2: ASIP Partial Specification

AE Tier – GMRobot Specification

At the AE tier, we specify the AEs - *GMRobot*, *Robot1*, *Robot2* and *Robot3*. In Listing 3, we present a part of the *GMRobot* specification, which describes the *GMRobots's* self-scheduling policy specification. To specify the self-scheduling policy, we use four major elements:

- *AE-level metrics* – AE metrics expressing the scheduling task deadline - *groupTaskDeadline*, the current scheduling task time - *groupTaskTime*, and the current number of idle robots in the group - *numberOfLocalIdleRobots*;
- *SELF_SCHEDULING* – the self-scheduling policy of the *GMRobot*. We use a set of *fluents* and *mappings* to specify the policy [3, 4]. With the fluents, we express specific situations in which the self-scheduling policy is interested, and with the mappings we map those situations to actions (see Listing 3);
- *actions* – a set of actions that could be undertaken by the *GMRobot* in response to certain conditions, and according to the self-scheduling policy;
- *events* – a set of events that are triggered by, the actions, according to the self-scheduling policies.

The following elements describe the specification in Listing 3.

```
AE GMRobot {
   AESELF_MANAGEMENT {
      OTHER_POLICIES {
         SELF_SCHEDULING {
            FLUENT inGroupTaskReceived { INITIATES: EVENTS.groupTaskReceived;
               TERMINATES: EVENTS.groupTaskScheduled
                           or EVENTS.groupTaskDenied }
            FLUENT inTaskImpossible { INITIATES: EVENTS.taskImpossibleReceived;
               TERMINATES: EVENTS.schedulingTaskAssigned
                           or EVENTS.lackOfIdleRobots }
            FLUENT inLackOfIdleRobots{ INITIATES: EVENTS.lackOfIdleRobots;
               TERMINATES: EVENTS.schedulingTaskAssigned or
                           EVENTS.groupTaskDenied}
            FLUENT inFinalizingGroupTask {
               INITIATES: EVENTS.taskCompletedWCETReceived or
                          EVENTS.taskCompletedECETReceived;
               TERMINATES: EVENTS.groupTaskDone or
                           EVENTS.groupTaskDenied }
            MAPPING { CONDITION: inGroupTaskReceived;
               ACTION: ACTIONS.setTaskDeadline;
               ACTION: ACTIONS.modelGroupTask }
            MAPPING {.... };   MAPPING {.... };   MAPPING {.... }
         } } // OTHER_POLICIES   } // AESELF_MANAGEMENT
   METRICS {
      METRIC groupTaskDeadline { ...}
      METRIC groupTaskTime { ...}
      METRIC numberOfLocalIdleRobots { ....}}
   ACTIONS {  ACTION setTaskDeadline {  ....
         DOES {  set METRICS.groupTaskTime = 0;
                 set METRICS.groupTaskDeadline =
                     ASIP.MESSAGES.groupTask.PARAMS.deadline;  }
      ONERR_TRIGGERS: EVENTS.groupTaskDenied;      }
      ACTION CallScheduleIndividualTask {
         PARAMETER: int taskID;
         RETURN_TYPE: int;
         DOES {  return call IMPL ScheduleIndividualTask (taskID) }
         ONERR_TRIGGERS: EVENTS.groupTaskDenied      }
      ACTION modelGroupTask {
         GUARDS: METRICS.groupTaskTime<=AS.METRICS.groupTaskDeadline;
         ENSURES: EVENTS.groupTaskScheduled;
         ENSURES: METRICS.groupTaskDeadline;
         DOES {  int taskID;
            do {  taskID = call ACTIONS.CallScheduleIndividualTask(
                  MESSAGES.groupTask.PARAMS.idTP);
                  IF taskID > 0 THEN call assignTaskToLocalIdleRobot(taskID);
            } while (taskID != 0)      }
         TRIGGERS: EVENTS.groupTaskScheduled;
         ONERR_TRIGGERS: EVENTS. groupTaskDenied;      }
      ACTION assignTaskToLocalIdleRobot {
         PARAMETER: int taskID;
         GUARDS:  METRICS.numberOfLocalIdleRobots > 0;
         ENSURES: EVENTS.schedulingTaskAssigned;
         DOES {  set ASIP.MESSAGES.requestScheduling.RECEIVER =
                     call IMPL AllocateLocalIdleRobot;
            set ASIP.MESSAGES.requestScheduling.PARAMS.idTP = taskID;
            set ASIP.MESSAGES.requestScheduling.PARAMS.deadline =
                METRICS.groupTaskDeadline - METRICS.groupTaskTime;
            call ASIP.FUNCTIONS.sendRequestSchedulingMsg;              }
         TRIGGERS: EVENTS. schedulingTaskAssigned;
         ONERR_TRIGGERS: EVENTS.lackOfIdleRobots;      }
      ACTION assignTaskToRemoteIdleRobot { ....
         DOES {  set ASIP.MESSAGES.requestScheduling.PARAMS.idTP =
                     call IMPL AllocateRemoteIdleRobot;
            call ASIP.FUNCTIONS.sendRequestSchedulingMsg;              }
         TRIGGERS: EVENTS.schedulingTaskAssigned;
         ONERR_TRIGGERS: EVENTS.groupTaskDenied;                    }
      ACTION finalizeGroupTask { ....
         DOES {  foreach member in AS.GROUPS.groupRobots.MEMBERS {
                 call ASIP.FUNCTIONS.receiveTaskCompleted;  }              }
         TRIGGERS: EVENTS.groupTaskDone;
         ONERR_TRIGGERS: EVENTS.groupTaskDenied;      }              }
   EVENTS { // these events are used in the fluents' specification
      EVENT groupTaskReceived:  received ASIP.MESSAGES.groupTask;
      EVENT taskImpossibleReceived:  received ASIP.MESSAGES.taskImpossible;
      EVENT schedulingTaskAssigned;
      EVENT lackOfIdleRobots;
      EVENT groupTaskDenied;
      EVENT groupTaskDone;
      EVENT taskCompletedWCETReceived:received taskCompletedWCET;
      EVENT taskCompletedECETReceived: received taskCompletedECET;
   } // EVENTS   }
```

Listing 3: AE GMRobot Partial Specification

inGroupTaskReceived. This fluent takes place when the *GMRobot* has received a group task-requesting message, thus sent by the *SuperiorRobot*. The fluent is initiated by the *groupTaskReceived* event, which occurs when the *GMRobot* has received a *groupTask* message. Moreover, this fluent terminates when the *groupTaskScheduled* event occurs, i.e. when the *GMRobot* has successfully modeled the group task,

scheduled the individual tasks, and sent them to the robots. The fluent can also terminates when the *groupTaskDenied* event occurs, i.e. when the deadline cannot be held. Further, this fluent is mapped to the *setTaskDeadline* action and to the *modelGroupTask* action. The former action initializes the *groupTaskTime* metric and sets the *groupTaskDeadline* metric. The latter action will perform while the time for the group task does not exceed the deadline (see the *GUARDS* clause in the action's specification), in which case a *taskDenied* event is fired. The *modelGroupTask* action loops:

- First, it calls the *CallScheduleIndividualTask* action to schedule from the group task individual tasks. The action uses the *ScheduleIndividualTask* **IMPL** routine [3] to do the scheduling. This routine returns the ID of a scheduled individual task, or zero when the entire group task has been scheduled. If the scheduling cannot be done, it fires the *groupTaskDenied* event.
- Second, it calls the *assignTaskToLocalIdleRobot* action to allocate to each individual task an idle robot. This action cannot be performed if there are no idle robots in the group (see the *GUARDS* clause in the action's specification), in which case a *lackOfIdleRobots* event is fired. The action uses the *AllocateLocalIdleRobot* **IMPL** routine (see the ASSL clause **IMPL** in [3, 4]) to allocate an idle robot. The **IMPL** clause states for "further implementation", which means that the ASSL framework will generate an empty routine for manual implementation.

inTaskImpossible. This fluent takes place when one of the robots cannot perform its task. The fluent is initiated by the *taskImpossibleReceived* event, which occurs when the *GMRobot* has received a *taskImpossible* message from a robot. Moreover, this fluent terminates when the *schedulingTaskAssigned* event occurs, i.e. when the *GMRobot* has successfully assigned the task to a new robot from the group. The fluent can also be terminated by the *lackOfIdleRobots* event, i.e. when there are no more idle robots in the group, and hence the task cannot be assigned. Further, this fluent is mapped to the *assignTaskToLocalIdleRobot* action, which allocates an idle robot, if any, and sends a *requestScheduling* message to it.

inLackOfIdleRobots. This fluent is triggered when the *lackOfIdleRobots* event occurs and terminates when the *schedulingTaskAssigned* event occurs, i.e. when the *GMRobot* has successfully assigned the task to a newly assigned to the group robot. The fluent can also be terminated by the *groupTaskDenied* event, i.e. when there are no more idle robots in the entire system, and hence the task cannot be performed, thus leading to a failure of the entire group scheduling task. Further, this fluent is mapped to the *assignTaskToRemoteIdleRobot* action, which is similar to the *assignTaskToLocalIdleRobot* action, but attempts to assign the task to an idle outside robot.

inFinalizingGroupTask. This fluent is triggered when the *GMRobot* has received a *taskCompletedECET* or a *taskCompletedWCET* message from the robots. The fluent terminates when of the events *groupTaskDone* and *groupTaskDenied* occurs. Further, the fluent is mapped to the *finalizeGroupTask* action. In this action, *GMRobot* waits for a "task completed" message to be received from all the robots. In case, not all the messages arrived in time, i.e. the group task time runs out the task deadline time (see the *GMRobot* metrics *groupTaskTime* and *groupTaskDeadline*) the action terminates with a *groupTaskDenied* event. If all the messages arrived in time, the action terminates with a *groupTaskDone* event.

AE Tier – Robot Specification

At this tier, we specify an AE robot class (AECLASS in ASSL). This class specifies three robot AEs in common: *Robot1*, *Robot2* and *Robot3*, which extend the *Robot* class (see the end of Listing 4). To specify the appropriate self-scheduling behavior we use a set of AE-level metrics to measure and observe:

- *timeDeadline* – the local scheduling task deadline;
- *taskTime* – measures the current task execution time;
- *timeECET/timeWCET* - the *ECET*/*WCET* scheduling task execution times.

```
AECLASS Robot {
   AESELF_MANAGEMENT {
      OTHER_POLICIES {
         TAFT_SELF_SCHEDULING {
            FLUENT inSchedulingRequest { // process the scheduling request
               INITIATES: EVENTS.schedulingRequestReceived;
      TERMINATES: EVENTS.taskTimesComputed  or EVENTS.taskImpossible }
            FLUENT inTaskImpossible {
               INITIATES: EVENTS.taskImpossible;
               TERMINATES: EVENTS.taskImpossibleSent }
            FLUENT inTaskMainPart{
               INITIATES: EVENTS.taskTimesComputed;
      TERMINATES: EVENTS.taskCompletedForECET  or EVENTS.ECETOut }
            FLUENT inTaskExceptionPart{
               INITIATES: EVENTS.ECETOut;
      TERMINATES: EVENTS.taskCompletedForWCET or EVENTS.taskImpossible }
            MAPPING {  CONDITION: inSchedulingRequest;
               ACTION: ACTIONS.computeTaskTimes }
            MAPPING { ... }  MAPPING { ... } MAPPING { ... }
      } } } // AESELF_MANAGEMENT
   METRICS {   METRIC timeDeadline { ... } METRIC timeWCET { ... }
      METRIC timeECET { ...} METRIC taskTime { ...} METRIC robotPosition {....}  }
   ACTIONS {   ACTION computeTaskTimes {
         ENSURES: METRICS.timeECET <= METRICS.timeWCET;
         ENSURES: METRICS.timeWCET <= METRICS.timeDeadline;
      ENSURES: METRICS.timeECET and METRICS.timeWCET and
                  METRICS.timeDeadline; // must be held at the end
         DOES { set METRICS.timeDeadline =
                     ASIP.MESSAGES.requestScheduling.PARAMS.deadline;
                  set Metrics.timeECET = call IMPL ComputeECETTime;
                  set Metrics.timeWCET = call IMPL ComputeWCETTime;    }
         TRIGGERS: EVENTS.taskTimesComputed;
         ONERR_TRIGGERS: EVENTS.taskImpossible; }
      ACTION runSchedulingTaskMainPart { ....
         DOES { call IMPL MainPart;
            call ASIP.FUNCTIONS.sendTaskCompletedECETMsg;
            call ASIP.FUNCTIONS.sendRobotPositionMsg; }
         TRIGGERS: EVENTS.taskCompletedForECET;
         ONERR_TRIGGERS: EVENTS.ECETOut; }
      ACTION runSchedulingTaskExceptionPart { ....
         DOES { call IMPL ExceptionPart;
            call ASIP.FUNCTIONS.sendTaskCompletedWCETMsg;
            call ASIP.FUNCTIONS.sendRobotPositionMsg;     }
         TRIGGERS: EVENTS.taskCompletedForWCET;
         ONERR_TRIGGERS: EVENTS.taskImpossible;          }
      ACTION notifyForSchedulingTaskFailure { ....
         DOES { call ASIP.FUNCTIONS.sendTaskImpossibleMsg;   }
         TRIGGERS: EVENTS. taskImpossibleSent;} } // ACTIONS
   EVENTS { // these events are used in the fluents' specification
      EVENT timeOut: not METRICS.taskTime;
      EVENT schedulingRequestReceived;
      EVENT taskTimesComputed;
      EVENT taskCompletedForECET; EVENT taskCompletedForWCET;
      EVENT taskImpossible;   EVENT taskImpossibleSent;
      EVENT ECETOut;
   } // EVENTS  } // AE Robot
// specify three similar robots by extending the Robot class
AE Robot1 extends Robot {};
AE Robot2 extends Robot {};
AE Robot3 extends Robot {};
```

Listing 4: AE Robot Partial Specification

As in the GMRobot specification, we specify the self-scheduling policy as an ASSL policy structure, called *TAFT_SELF_SCHEDULING*. The specification of this policy is based on the TAFT [6] scheduling algorithm (see Section III):

inSchedulingRequest. With this fluent we process the scheduling request from the GMRobot. The fluent starts when the message requesting for the execution of the scheduling task has been received (see the event *schedulingRequestReceived in Listing 4*), and terminates when the *ECET* and *WCET* task execution times have been successfully computed (see the event *taskTimesComputed* in Listing 4). This fluent could also terminate with the event *taskImpossible*, which will occur if the *ECET* and *WECT* task execution times cannot be computed, or the *ECET* is bigger than the *WCET*, or the deadline is less than the *WCET*. All this is specified in the *computeTaskTimes* action, which is mapped to this fluent. This action uses two **IMPL** routines to compute the *WCET* and *ECET* execution times respectively.

inTaskImpossible. With this fluent we handle the situation when scheduling is not possible, due to a failure to compute the *ECET* and *WECT* task times (see the *computeTaskTimes* action*)* or a failure to execute the *ExceptionPart* of the scheduling task (see the *runSchedulingTaskExceptionPart* action). The fluent starts with the *taskImpossible* event and terminates with the *taskImpossibleSent* event. The latter occurs when the message *taskImposiible* has been sent to the GMRobot. This is specified in the *notifyForSchedulingTaskFailure* action.

inTaskMainPart. With this fluent we process the execution of the MP of the scheduling task. The fluent starts with the *taskTimesComputed* event, which is triggered when the *ECET* and *WECT* task times are computed successfully (see the *computeTaskTimes* action). Moreover, the fluent terminates with the *taskCompletedForECET* event, which is triggered by the *runSchedulingTaskMainPart* action. The latter is mapped to this fluent, and calls an **IMPL** routine to execute the MP part of the scheduling task. Upon successful completion of **IMPL** routine, the action sends two messages to the GMRobot – one for successful completion of the task within the *ECET* time and one with the current robot position. In case the current execution time runs out the *ECET* (the *timeECET* metric), the action triggers the *ECETOut* event that also terminates the *inTaskMainPart* fluent.

inTaskExceptionPart. With this fluent we process the execution of the EP of the scheduling task. The fluent specifies the situation when the MP of the scheduling task cannot be executed due to timeout (see the *ECETOut* event), and the robot switches to the EP part of the task. This fluent is mapped to the *runSchedulingTaskExceptionPart* action, which calls an **IMPL** routine to execute the EP part of the scheduling task. Upon successful completion of **IMPL** routine, the action sends two messages to the GMRobot – one for successful completion of the task within the *WCET* time and one with the current robot position. In case the current execution time runs out the deadline time (measured by the *timeDeadline* metric), the action triggers the *taskImpossible* event.

V. Related Work

The problem of scheduling local processing tasks in Team Robotics, where all tasks must be performed reliably and in real time, has been tackled (see, for instance, [6, 7, 8]). In [8], a mechanism for scheduling motion control tasks in a multi-robot mobile embedded system is proposed in which the robots fuse the data obtained by their laser scanners to achieve more complete and accurate detection, localization and categorization of the objects in their environment. In [7], a scheduling algorithm is presented to guarantee the timely execution of critical tasks during overload situations. We have chosen to formalize task modeling and the underlying scheduling mechanism (Time-Aware Fault-Tolerant (TAFT)) scheduler described in [6]. Our decision is based of the fact that, while [7, 8] have already presented the basic ideas of the TAFT scheduling strategy, the algorithm proposed in [6] details the scheduling strategy adopted to achieve efficient runtime scheduling.

VI. Conclusion and Future Work

One of the most important aspects of Team-Robotics is scheduling, a feature requiring a formal mechanism for the self-scheduling and monitoring of tasks in a team of supervised autonomic mobile robots at runtime. In this paper, a self-scheduling specification and monitoring self-control method and a means to diagnose potential behavioral flaws, and consequently scheduling problems in Team Robotics AS-TRM architecture refined and specified formally with ASSL, is introduced. This paper reported on the initial steps toward the achievement of the above goal, namely:

1) Propose an AS-TRM layered architecture to a team of reactive, autonomic and synchronously communicating robots.

2) Describe a self-scheduling mechanism and introduce the model of a schedule as a labeled transition system.

3) Formalize the self-scheduling design and monitoring mechanism in Team Robotics with ASSL [3].

Further research is required to benchmark the approach in real-world system environments in order to ensure the correctness of the self-scheduling and monitoring mechanism in an evolving system Our future research directions include: (1) runtime management of the requirements (the exact timing requirements may not be fully known until the robot starts operating in a potentially unknown, dynamic and often unpredictable environment); (2) incorporating task requirements on priority, periodicity and non-preemptiveness (when other tasks cannot be executed until the task is completed). We must develop and analyze algorithms and negotiation protocols for conflicting timing and scheduling requirements, and determine what bidding or negotiation algorithms are the most effective. These are some of the issues for future investigation.

References

[1] IBM Corporation, “An architectural blueprint for autonomic computing”, White Paper, 4th Edition, 2006.

[2] E. Vassev, H. Kuang, O. Ormandjieva and J. Paquet, "Reactive, Distributed and Autonomic Computing Aspects of AS-TRM", *Proceedings of 1st International Conference on Software and Data Technologies - ICSOFT'06*, pp. 196-202, 2006.

[3] E. Vassev and J. Paquet, "ASSL - Autonomic System Specification Language", *Proceedings of the 31st Annual IEEE/NASA Software Engineering Workshop (SEW-31)*, pp. 300-309, Baltimore, MD, USA, 2007.

[4] E. Vassev and J. Paquet, Towards an Autonomic Element Architecture for ASSL, *Proceedings of the 29th International Conference on Software Engineering / Software Engineering for Adaptive and Self-managing Systems (ICSE 2007 SEAMS)*, p. 4, Minneapolis, MN, USA, 2007.

[5] S.W. Cheng, D. Garlan and B. Schmerl. "Architecture-based Self-adaptation in the Presence of Multiple Objectives", *Proceedings of the ICSE 2006 Workshop on Software Engineering for Adaptive and Self-Managing Systems (SEAMS)*, 2006.

[6] L.B. Becker, E. Nett, D. Dchemmer and M. Gergeleit. "Robust Scheduling in Team-Robotics", *Journal of Systems and Software 77 (2005)*, pp. 3-26, 2005

[7] P. Richardson, L. Sieh, A. Elkateeb. "Fault-Tolerant Adaptive Scheduling for Embedded Real-Time Systems", *IEEE Micro*, pp. 41-51, Sept-Oct 2001.

[8] S. Schemmer, E. Nett, M. Mock. "Reliable Real-Time Cooperation of Mobile Autonomous Systems", *Proceedings of SRDS*, pp.238-246, 2001.

[9] O. Ormandjieva, H. Kuang, E. Vassev. "Reliability Self-Assessment in Reactive Autonomic Systems: Autonomic System - Time Reactive Model Approach". *International Transactions on Systems Science and Applications, Volume 2 (1)*, pp. 99-104, 2006.

A Novel Control System for a Direct Drive Linear Permanent Magnet Actuator with Intrinsic Position Hold

[1]Evgueni Sliva, *Member, IEEE,* [1]Bijan Sobhi-Najafabadi, [1]Michael Tonkin, [1]Robert Speedie and [2]Sid Sobhani

[1]Centre for Advanced Manufacturing Research, University of South Australia.
Mawson Lakes Boulevard
Mawson Lakes 5095 SA
Australia
[2]Schefenacker Vision Systems Australia Pty Ltd.
Sherriffs Road, Lonsdale, 5160 SA Australia

***Abstract*-A novel control and mechanical system design of a linear direct drive permanent magnet actuator is presented. The actuator consists of a Halbach permanent magnet array for the moving part and a two-phase air core stator winding. The magnet assembly is clamped to the stator with springs that maintain the actuator in a fixed position when no power is applied. By applying a sequence of controlled current pulses both separating and propelling forces are generated. A prototype of the actuator was built. The control system was developed and tested with the actuator prototype. The proposed control algorithms require neither high computational capacity of the hardware nor high resolution velocity information. A model was developed to assess the control system performance. The control system design and associated test results are discussed.**

I. INTRODUCTION

There is an increasing demand for permanent magnet linear actuators. These actuators were conventionally realized using a rotating motor and a mechanical arrangement to convert rotary into linear motion. An example of such a linear actuator is presented [1]. This actuator uses a brushless DC motor with a compact planetary roller spindle drive, which converts rotary motion of the motor into the linear displacement of the output shaft. For better control this actuator is equipped with force, linear position and velocity sensors. Similar but simpler actuators can be used for less demanding applications, eg for automotive applications such as rear view mirrors and seat position actuators. The great mechanical advantage of these actuators allows them to use small volume high speed electrical motors [2]; these actuators can also brake the motion of the output shaft when the actuator power is switched off, which is very desirable for these applications.

Compared to conventional actuators, direct drive permanent magnet actuators offer advantages such as replacing gears, increasing movement control and widening the mechanical bandwidth and lowering maintenance. Gearless permanent magnet linear actuators, competing with other type of actuators, are used in applications where their advantages are most beneficial such as aerospace, transportation, industrial automation, biomedical systems [3], [4], [5].

The main difficulty for direct drives is to achieve sufficient force density, which is force-to-mass ratio [6]. High force density can be can be achieved by using rare earth magnets [7]. A Halbach permanent magnet array offers a number of attractive features as they eliminate the need for a back iron thereby reducing the mass and inertia of the actuator [4], [3]. These magnets can be manufactured relatively easily using injection molded or sintered magnets [3].These magnets can be manufactured in many geometries to achieve rotary, linear, two and three degree of freedom motion [4]. An actuator with a Halbach magnet can independently control orthogonal forces, which was used for the planar actuator with electromagnetic levitation [8].

The dynamics of these actuators can be is described by essentially linear equations so the traditional control approach using PID governor can be used [4]. Their dynamic models are similar to the permanent magnet-coil actuators and the control approach used for them can also be used [9]. An appropriate bandwidth velocity and position feedback is needed to implement this control which can affect the viability of these actuators for cost critical applications.

Direct drive geared actuators ability to hold the position of the shaft when the power is off, is often an essential feature. This problem was solved in the friction drive magnetic motor, which has a magnet mover with inclined spring legs and a solenoid coil wound around the tubular shaft [10]. This motor uses resonant excitation of the mechanical system and its performance is very frequency sensitive.

Automotive actuators do not need high precision compared to the mentioned application areas for direct drive actuators, and are used to move an object from one position to another with an ergonomic steady average velocity. They need to hold the position when the power is off.

This paper addresses these problems for a direct drive permanent magnet linear actuator. A novel control algorithm for Halbach permanent magnet synchronous linear actuators is proposed. This algorithm uses the ability of the Halbach permanent magnet actuator to produce force not only in the lateral direction but also in the vertical direction, which is used in a friction lock. A permanent magnet linear actuator mechanical system was modified and a control algorithm implemented to prove the concept of such an actuator.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 77–82.

II. CONSTRUCTION OF AN ELECTROMAGNETIC ACTUATOR

Fig. 1 shows the cross section of the proof-of-concept two phase Halbach permanent magnet linear actuator, which was designed and prototyped. A photo of the actuator prototype is presented in Fig. 2. A stationary part has a base with the two-phase coil set up removed. On top of the base a frame

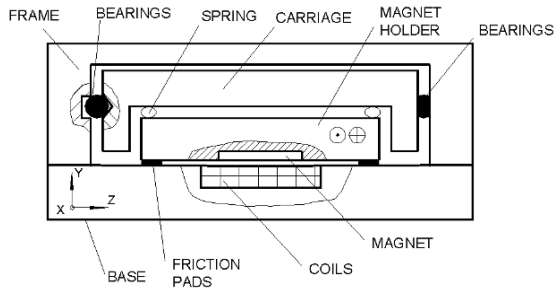


Fig. 1. Cross section diagram of the actuator prototype

was mounted. The frame has linear bearings in which a carriage is supported. This is achieved with ball bearings on straight grooves cut on the frame and the block. A NdFeB plastic composite bar is magnetized as a Halbach array and

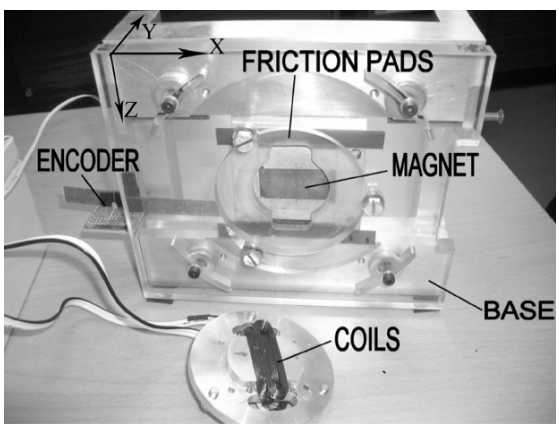


Fig. 2. Actuator prototype with the coil extracted from the assembly

it is mounted on a nonmagnetic block (magnet holder) which in turn is attached to the carriage by two tubular springs. The magnet is 35 mm long, 18 mm wide and 3 mm high. Tubular springs are made of polyethylene 6 mm diameter and 0.1 mm wall thickness. The mass of the magnet assembled with the holder was 0.035 kg and the mass of the carriage was 0.120 kg. Friction pads are mounted on the carriage and on the base as shown in the figure. When the friction pads were engaged the gap between the surfaces of the coils and the magnet was 0.2 mm. The air core coils had 60 turns each of 0.33 mm laminated copper wire with an inductance of 230 μH and resistance of 2.4 Ohm. The displacement of the carriage was measured by an optical linear encoder with the resolution of 0.18 mm. The encoder was connected to the frame and the encoder scale attached to the carriage.

The actuator works as follows. When the power is assembly is kept stationary by the friction force. The motion is achieved by injecting a series of current pulses into the stator coils, creating vertical and lateral forces which are applied to the magnet. The vertical force pulse compresses the tubular springs and disengages the friction pads so the carriage can move freely in the X direction while elevated. Then the second force pulse is applied in ○←X direction depending on the direction required. The pulse frequency controls the mean velocity of the mover. Details of the development of the control algorithms for this actuator are presented in the subsequent sections.

III. CONTROL SYSTEM HARDWARE

A control system hardware diagram is presented in Fig. 3. A development board eZdsp F2812 was used as a core for the control hardware. This board uses a 32-bit high-performance 150 MHz (6.67-ns Cycle Time) DSP controller from Texas Instruments. This controller has on board a broad range of functional blocks including timers with 6 PWM outputs and two encoder signal processing units. The outputs of the board are connected to the power amplifiers U5, U6 providing PWM and sine signals for controlling the switching bridge ICs. These ICs also supply current feedback signals (R1, R2), into the stator coils (L1, L2). The inductance and resistance of each of the coils was measured by a PEAC LCR meter as 230 μH and 2.4 Ohm respectively. The frequency of the PWM was set to 100 kHz. The switching amplifier power supply was 50 V. Current control was implemented with the DSP in a loop. The position feedback encoder is connected to the capture inputs of the DSP. The board has an embedded JTAG adapter for programming and debugging the DSP, which is connected to the computer via printer port EPP. A serial interface RS232 is used to facilitate data exchange with the main computer and is used for data and constants transfer to and from the board. This feature was used in automated tests when the computer running Matlab transferred the set of the parameters to the controller, initializing the test and acquiring the data read by the ADC card.

IV. MODEL OF MAGNET-COIL INTERACTION

The interaction of the coil currents with the magnetic field produces force and consequently the actuator motion. The amplitude and the direction of the force depend on coil currents and the permanent magnet flux, which is dependant upon the magnet assembly position. The equations for the generated force are given similarly to [8]. These equations were used to achieve independent control of vertical and lateral forces via appropriate current control. Impulsive control is used to produce motion. Control algorithms were developed, implemented and validated using appropriate hardware. For the actuator control it is crucial to characterize the interaction between the coils and magnet.

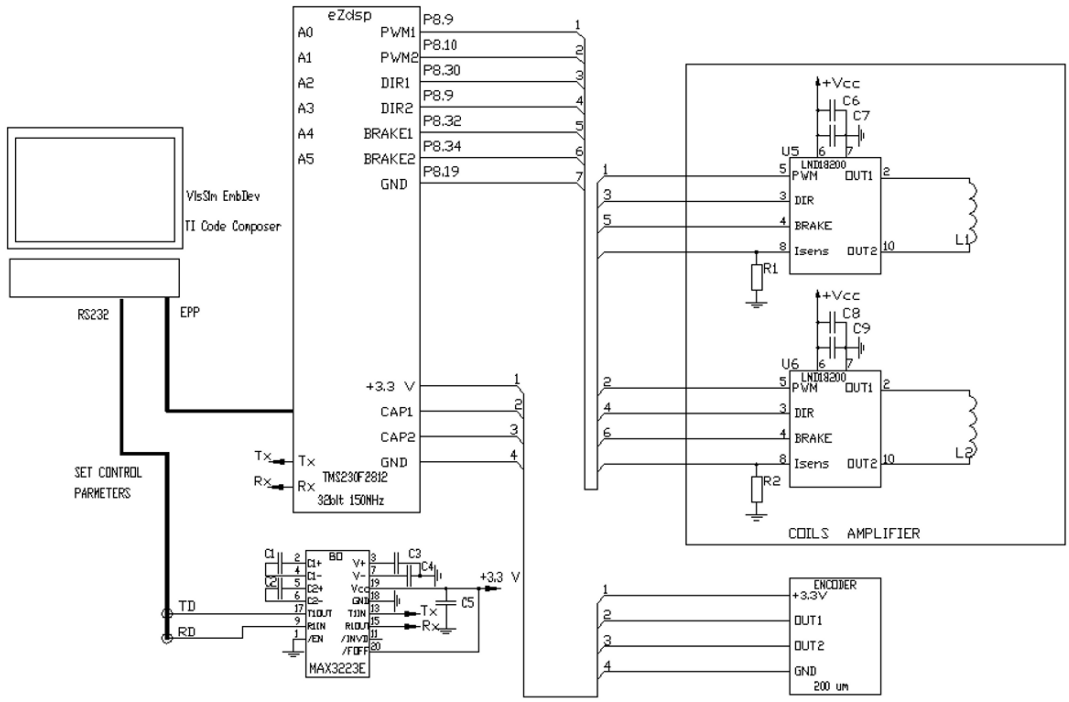


Fig. 3. Control system hardware diagram

The novel control algorithm presented in this paper is based on the amplitude and direction controlled force vector generation through the interaction of the permanent magnet and the stator coils. A modified Halbach permanent magnet synchronous actuator was used. A key element of this

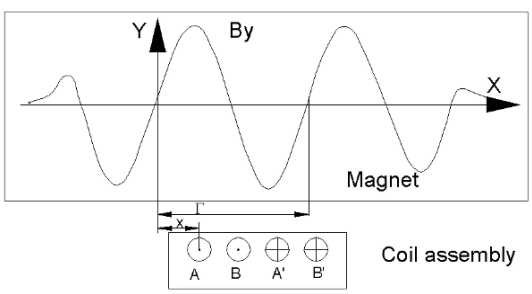


Fig. 4. Schematic of magnet-coil interaction

actuator is the rare earth permanent magnet with the magnetic flux rotating along the length of the magnet, which gives sinusoidal normal B_z and longitudinal B_x flux components [11], [12]. The Halbach magnet can be implemented in various geometries such as bar, ring or tubular. This actuator has a bar magnet.

The actuation is achieved by the interaction between the coil current and the magnet (Fig. 4). The normal flux component B_y is shown on the diagram. The coils are represented by the forward *(A,B)* and return *(A',B')* wires with the equivalent magnetomotive force of ni_A, ni_B and the active length of the coils *l* which is the same for both of the coils. The control system calculates and establishes the coil currents i_A, i_B using the instant position *x*r. The resulting lateral and vertical components of the force vector $\boldsymbol{F}(x) = [F_x(x)\ F_y(x)]^T$ can be calculated as

$$F(x) = nl\begin{bmatrix} B_{yA} - B_{yA'} & B_{yB} - B_{yB'} \\ B_{xA} - B_{xA'} & B_{xB} - B_{xB'} \end{bmatrix}\begin{bmatrix} i_A(x) \\ i_B(x) \end{bmatrix} \quad (1)$$

The flux density vector $B = [Bz\ Bx]^T$ ψalong the *x* axis of the magnet in this model case is changing as

$$B(x) = B_{\max}\sin\begin{bmatrix} \omega(x) \\ \omega(x) + \dfrac{\pi}{2} \end{bmatrix} \quad (2)$$

where the cyclic frequency of the flux density distribution is $\dfrac{2\pi}{\Gamma}$ (Fig. 4) . The origin of the axis *x* is assigned to the position with a zero vertical flux distribution $B_Y = 0$. The coils are designed so that the distance between their elements is equivalent to a quarter of the magnetic flux period Γ. The field distribution through the coil elements $\mathrm{A}, \mathrm{B}, \mathrm{A}', \mathrm{B}'$ with the coordinates of x, $\mathrm{x} + \dfrac{\pi}{2\omega}$, $\mathrm{x} + \dfrac{\pi}{\omega}$, $\mathrm{x} + \dfrac{3\pi}{2\omega}$ has a phase shift of the period quarter so:

$$B_{yA} = -B_{yA'} = B_{\max}\sin\omega x$$

$$B_{yB} = -B_{yB'} = B_{\max}\sin(\omega x + \frac{\pi}{2})$$

$$B_{xA} = -B_{xA'} = B_{\max}\sin(\omega x + \frac{\pi}{2})$$

$$B_{xB} = -B_{xB'} = -B_{\max}\sin(\omega x + \frac{\pi}{2})$$

ψ

As the result Eqn. 1 can be rewritten as follows

$$F(x) = 2B_{\max} nl \sin\begin{bmatrix} \omega x & \omega x + \frac{\pi}{2} \\ \omega x + \frac{\pi}{2} & -\omega x \end{bmatrix}\begin{bmatrix} i_A(x) \\ i_B(x) \end{bmatrix} \quad (3)$$

To produce just a lateral force the control system supplies into the coils a position dependant quadrature current with an amplitude of I_x:

$$\begin{bmatrix} i_A(x) \\ i_B(x) \end{bmatrix} = I_x \sin\begin{bmatrix} \omega x \\ \omega x + \frac{\pi}{2} \end{bmatrix} \quad (4)$$

Thus ←

$$F_x(x) = 2B_{\max}\sin\begin{bmatrix} \omega x \\ \omega x + \frac{\pi}{2} \end{bmatrix}^T I_x \sin\begin{bmatrix} \omega x \\ \omega x + \frac{\pi}{2} \end{bmatrix} \quad (5)$$

Which, when simplified, gives

$$F_x(x) = 2lB_{\max} I_x \quad (6)$$

Similarly a vertical force is produced when the control system supplies current in quadrature according to:

$$\begin{bmatrix} i_A(x) \\ i_B(x) \end{bmatrix} = I_y \sin\begin{bmatrix} \omega x + \frac{\pi}{2} \\ -\omega x \end{bmatrix} \quad (7)$$

hence

$$F_y(x) = 2lB_{\max} I_y \quad (8)$$

Equations Eqn. 6, Eqn. 8 show that both lateral and vertical components of the force vector can be controlled independently with the appropriate sinusoidal quadrature currents supplied to the coils as shown in Eqn.4 and Eqn. 7. These quadrature functions are implemented by the digital control system so that the internal variables I_x and I_y, of Eqn. 4 and Eqn. 7 with position information, x, from the encoder are used. The control system hardware is also equipped with a high frequency PWM digital to analogue converter and a switching amplifier. The dynamics and the discrete nature of the system is not considered here as they have a much wider bandwidth than the utility signals in the system.

With these assumptions the control system is capable of producing lateral F_x or vertical F_y forces proportional to the internal control system variables I_x, I_y with coefficients K_{FX}, K_{FY}, ideally equal to $2lB_{max}$.

In the following control system diagram the force projections F_x and F_y control are used as input variables.

This model of the force generated will be used in the actuator control development. The limitations of the model also include the assumption that the coils are not considered as distributed, which can be overcome by using FEM technique if needed.

Conventional synchronous permanent magnet motors are designed to produce only a lateral force from sinusoidal currents with the appropriate phase shift supplied to the coils [13],[11]. In the proposed actuator both lateral and normal components of the force vector are used as explained in the following section.

V. DEVELOPMENT OF CONTROL ALGORITHM

For impulsive control the motion is produced as a consequence of the steps produced by the vertical and lateral force pulses as shown in Fig. 5.

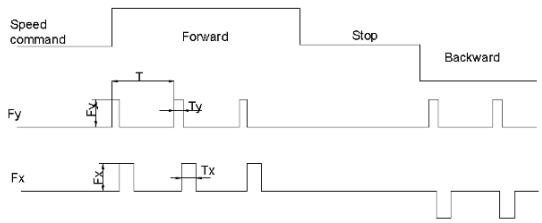


Figure 5. Rectangular force pulses control

The model of the mechanical system is presented in Fig. 6. The sequence of F_x, F_y pulses is generated by the electromechanical system of the actuator when the control system supplies appropriate currents i_A, i_B into the coils as was explained above. To make one step the friction force is released by applying a vertical force to compress the clamping springs and provide the rotor to stator clearance. Then the lateral force is applied to move the magnet. A model of the actuator was developed in Matlab Simulink toolbox. The sequence of pulses with appropriate timing and amplitude is preprogrammed to produce motion in the positive direction as shown in Fig. 5. Initially a vertical force pulse was produced. When this force overcame the *Clamping force* the mass is moved freely in a vertical direction and the gap between the friction pads is created.

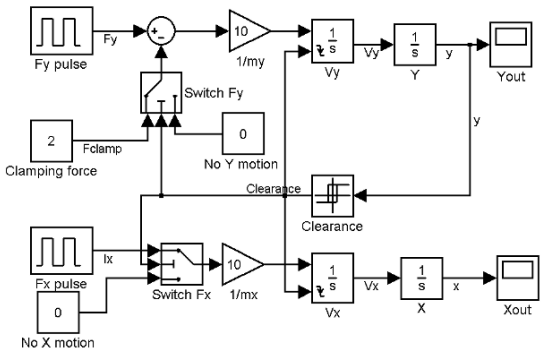


Figure 6. Diagram of mechanical model of the actuator.

During the time when there is a gap between the pads, the lateral force pulse F_x is produced. This force moves the sliding carriage freely until the pads engage again. The motion being possible when there is a gap between the stator and rotor. Damping or deformation of the stator and rotor bodies are not considered.

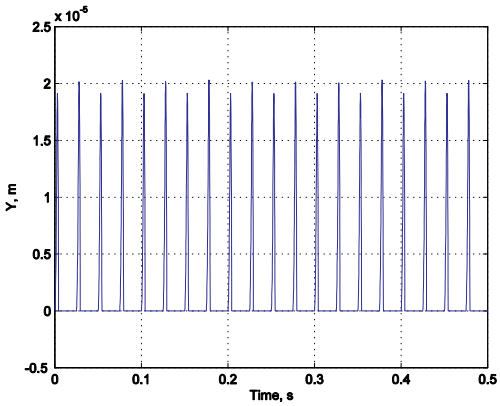


Figure 7. Simulation results of vertical displacement of the carriage.

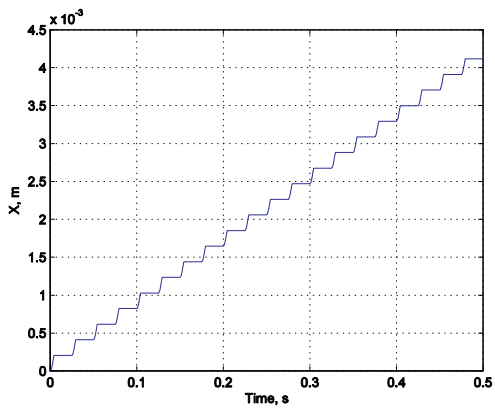


Figure 8. Simulation results of lateral displacement of the carriage.

The model was supplied with the previously measured mechanical parameters such as the clamping force, the mass of the moving part assembly etc. The force pulses were applied with the $Fy=2.5N$, $F_x=2N$, $T_y=1ms$, $T_x=1ms$, $T=20ms$. The results of the simulation are presented in Fig. 7 and Fig. 8. When the pulse of the vertical force is applied the rotor is accelerating up and then decelerating down to the stator. During this action there is clearance between the rotor and stator permitting motion and a lateral force impulse is applied. The rotor mass is accelerating during this pulse and then moving along a ballistic trajectory.

VI. CONTROL SYSTEM WITH ACTUATOR TESTS

An impulsive control system was implemented with the actuator prototype and tested. The force components F_y and F_x control were implemented according to Eqns. 6 and 8. The values of the current supplied into the coils are calculated before each motion step and are not altered during the pulse. As the duty cycle of the force pulses is very low there is significant time to process these calculations. The current to force conversion coefficients $K_{FX}=0.49\ N/A$, $K_{FY}=0.51\ N/A$ were measured. The variation of these coefficients along the actuator work range did not exceed 10%. These coefficients were used for presetting I_x and I_y to achieve appropriate force pulses. The friction force between the engaged pads was measured as 2 N.

After the control system was set up t tests of the actuator were run. When the system was loaded with the parameter used for the simulation Fig. 8 the actuator did not produce consistent motion. The results of the test with

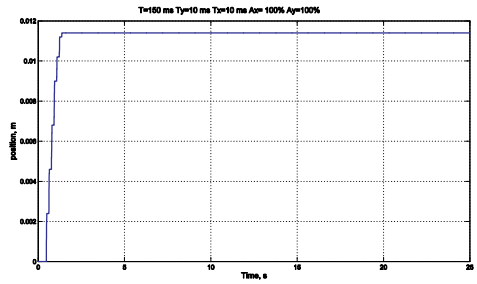


Figure 9. Step response of the impulsively controlled actuator.

$F_x = F_y = 2.7N$ and timing of $T_x = T_y = 10ms$, $T = 150ms$ are presented in Fig. 9. The step varies along the stroke from 0.5 mm to 2 mm, which can be due to the variation of the developed force, friction force and tilt issue. Acoustic noise was significant.

VII. CONCLUSIONS

A novel impulsive control algorithm for a permanent magnet actuator was proposed and implemented in the proof of concept prototype. The prototype holds the position when the power is switched off. The computational capacity of the algorithm is low and can be implemented with a simple microcontroller. This algorithm neither needs precise position information or computational capacity as the duty cycle of the pulses applied is usually less than 10% with the frequency as low as 25 Hz. The time between the pulses is long enough to accomplish this task with a simple microcontroller. The control system does not need to have a fast response or real time measurement of velocity. The dynamic range of the current amplifier is less than required for the case of the free-running prototype.

However there are some shortcomings such as the dynamic response of the system for the condition when the whole moving part is lifted and moved. This leads to substantial energy loses and narrows the bandwidth of the

system. The parameters of the prototypes and models were not optimized so the test results presented in this paper are useful only as preliminary data.

VIII. Acknowledgment

The work was carried out at the University of South Australia with support from the Cooperative Research Centre for Intelligent Manufacturing Systems and Technologies and Schefenacker Vision Systems Australia (SVSA).

References

[1] E. Kramer, H. Liu, N. Seitz, and G. Hirzinger, "A multisensory linear actuator system," *Mechatronics, IEEE/ASME Transactions on*, vol. 7, pp. 182 –185, 2002.

[2] B. Sobhi-Najafabadi, D. Halim, E. Sliva, and M. Tonkin, "Multiple-rotor permanent magnet dc motor topologies for low-cost manufacturing," *International Manufacturing Leaders Forum on Global Competitive Manufacturing*, 2005.

[3] Z. Zhu and D. Howe, "Halbach permanent magnet machines and applications: a review," *Electric Power Applications, IEE Proceedings-*, vol. 148, pp. 299 – 308, 2001.

[4] D. Howe, "Magnetic actuators," *Sensors and Actuators A: Physical*, vol. 81, pp. 268 –274, April 2000.

[5] J. F. Gieras and Z. J. Piech, *Linear Synchronous Motors: Transportation and Automation Systems*.1em plus 0.5em minus 0.4emBoca Raton, FL: CRC Press, 2000.

[6] P. Cavarec, H. Ben Ahmed, and B. Multon, "New multi-rod linear actuator for direct-drive, wide mechanical bandpass applications," *Industry Applications, IEEE Transactions on*, vol. 39, pp. 961–970, July-Aug. 2003.

[7] O. Ustun and R. Tuncay, "Design, analysis, and control of a novel linear actuator," *Industry Applications, IEEE Transactions on*, vol. 42, pp. 1007 – 1013, 2006.

[8] I. J. C. Compter, "Electro-dynamic planar motor," *Precision Engineering*, vol. 28, pp. 171 –180, April 2004.

[9] S. Liu, P. Mercorelli, and K. Lehmann, "Robust control of a permanent magnet linear motor in the presence of large disturbances," *Control Applications, 2003. CCA 2003. Proceedings of 2003 IEEE Conference on*, vol. 2, pp. 1301 –1306, 2003.

[10] T. Honda, N. Miyazaki, and J. Yamasaki, "Development of a magnetic linear motor based on a friction drive," *Magnetics, IEEE Transactions on*, vol. 35, pp. 4004 – 4006, 1999.

[11] Z. Zhu, P. Hor, D. Howe, and J. Rees-Jones, "Novel linear tubular brushless permanent magnet motor," *Electrical Machines and Drives, 1997 Eighth International Conference on (Conf. Publ. No. 444)*, pp. 91–95, 1997.

[12] J. Wang, D. Howe, and G. Jewell, "Analysis and design optimization of an improved axially magnetized tubular permanent-magnet machine," *Energy Conversion, IEEE Transactions on*, vol. 19, pp. 289 – 295, 2004.

[13] S. A. Nasar, I. Boldea, and L. E. Unnewehr, *Permanent Magnet, Reluctance and Self-Synchronous Motors*.Boca Raton, FL: CRC Press, 1

Characterizing the Exact Collision Course in the Plane for Mobile Robotics Application

K. Bendjilali[1], F. Belkhouche[2] and T. Jin[2]
[1]ECE Department, Lehigh University
[2]Texas A&M Intl University
fbelkhouche@tamiu.edu

Abstract- **In this paper, we discuss the exact collision course in a dynamic environment between a wheeled mobile robot and a moving object. The paths intersection conditions in the horizontal plane are deduced based on the geometry of the paths, and the collision course is deduced based on the relative kinematics model between the robot and the moving object. The exact conditions under which the robot and the moving object are in a collision course are derived and proven rigorously. The collision course condition is expressed as a function of the robot and the moving object orientation angles and linear velocities. The method can be used for collision detection in a dynamic environment, and therefore, it can be used for collision avoidance. Several simulation examples are used for an illustration.**

I. Introduction

Mobile robot navigation is among the most important topics in mobile robotics. The literature on robot navigation is substantial, where various methods and algorithms were suggested ([1],[2],[3],[4],[5]). Collision avoidance is another important topic highly connected to navigation. Clearly, collision detection is necessary for collision avoidance algorithms. Sensor-based collision avoidance methods are widely used in practice. In general collision detection with stationary objects can be performed easily using nowadays technology. The problem of collision detection becomes more difficult for moving objects. Similarly to navigation in static environment, navigation in dynamic environment has also seen important developments in the last decades. Methods based on different approaches such as dynamic vector field [6], path-velocity decomposition ([7], [8], [9]), vision [10], and other methods ([11], [12], [13]) were suggested. Navigation in dynamic environment requires the detection of the collision course between the robot and the moving object (or obstacle). Various techniques for collision detection are discussed in the literature. For example, neural networks were used for real time collision detection in [14]. Adaptive control law was suggested in [15] for dynamic collision detection. Hierarchies of detail based on spheres were also used in ([16], [17]). Hough transform for distance computation and collision avoidance was suggested in [18].

Our aim in this paper is to discuss the collision course between a robot and a moving object in the horizontal plane, where we use the relative kinematics equations between the robot and the moving object. The collision course conditions are derived in term of the angular and linear velocities, and the relative velocities. The robot is modeled as wheeled mobile robot of the unicycle type and the moving object is modeled as a geometric point. Extension to moving objects with certain particular shapes such as circular shape is straightforward.

II. Kinematics and Geometry

The world is attached to a global reference frame of coordinates {G}. The robot, denoted by R and a moving object denoted by B are moving in the horizontal plane according to the following kinematics equations

$$\begin{aligned} \dot{x}_R &= v_R \cos\theta_R \\ \dot{y}_R &= v_R \sin\theta_R \\ \dot{\theta}_R &= \omega_R \end{aligned} \qquad (1)$$

and

$$\begin{aligned} \dot{x}_B &= v_B \cos\theta_B \\ \dot{y}_B &= v_B \sin\theta_B \end{aligned} \qquad (2)$$

respectively. (x_R,y_R) represent the robot's position in the global reference frame of coordinates, (x_B,y_B) represents the moving object position in the global reference frame of coordinates. θ_R and θ_B are the robot's and the moving object's orientation angles with respect to the positive x-axis, respectively. v_R and ω_R are the linear and angular velocities of the robot, respectively. v_B is the moving object's linear velocity. No assumption is made on the motion of the moving object; this means that accelerating objects are not excluded. We consider the geometry of the navigation problem shown in figure 1. The line of sight LOS is the imaginary straight line starting from the robot reference point and directed towards the moving object. The relative distance between R and B is given by

$$r_{BR} = \sqrt{(x_B - x_R)^2 + (y_B - y_R)^2} \qquad (3)$$

We define the line of sight angle λ_{BR} as the angle from the positive x-axis to the line of sight. The expression for λ_{BR} is as follows

$$\tan\lambda_{BR} = \frac{y_B - y_R}{x_B - x_R} \qquad (4)$$

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 83–88.

Consider the relative velocity vector

$$\vec{v}_{BR} = \vec{v}_B - \vec{v}_R \qquad (5)$$

The relative velocity can be decomposed into two components in the Cartesian plane as follows

$$\vec{v}_{BR} = v_{BR}^x \vec{U}_x + v_{BR}^y \vec{U}_y \qquad (6)$$

where U_x and U_y are the unit vector along the x and y-axes, respectively. By using the robot's and the moving object kinematics equations, we get

$$\begin{aligned} v_{BR}^x &= \dot{x}_d = v_B \cos\theta_B - v_R \cos\theta_R \\ v_{BR}^y &= \dot{y}_d = v_B \sin\theta_B - v_R \sin\theta_R \end{aligned} \qquad (7)$$

with $x_d=x_B-x_R$, $y_d=y_B-y_R$. System (7) provides a complete description of the motion of the moving object seen by the robot. However, this system is highly nonlinear and difficult to solve analytically.

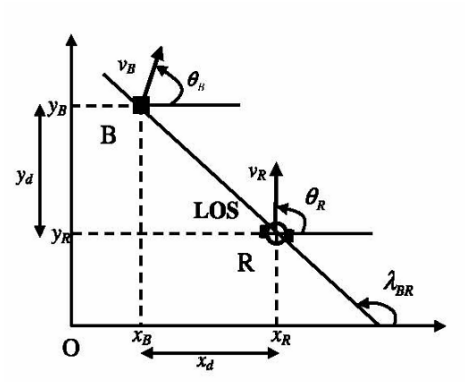


Fig. 1: Geometry of the navigation problem

III. COLLISION COURSE

Roughly speaking, two objects are in collision course if they are approaching from each other, i.e., $\dot{r} < 0$ at a given time or in a given time interval. However, this condition can lead to a wrong conclusion about collision. Consider the scenarios depicted in figure 2, where the robot and the moving object are moving at the same speed. The robot and the moving object are approaching from each other in the time interval $[t_0,t_3]$, but they are not in a collision course, since the robot reaches the paths intersection point before the moving object.

Our aim is to give an exact characterization of the collision course based on the relative kinematics equations. This will allow the prediction of future collision course based on the states of the robot and the moving object. These states are characterized by: v_R, v_B, θ_R, θ_B, λ_{BR}.

We define the paths intersection as follows:

Definition 1: Paths intersection means that the path traced by the robot and the moving object will intersect at a given point P.

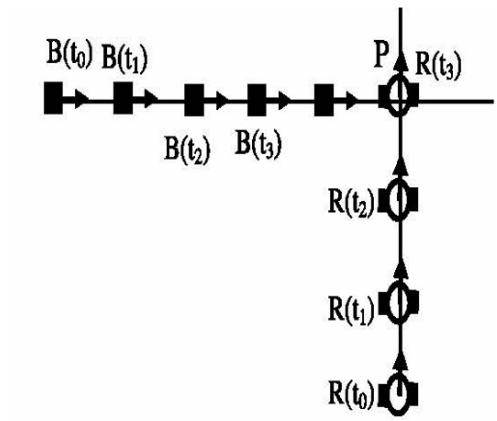


Fig. 2: An example where the robot and the moving object are approaching from each other but they are not in a collision course

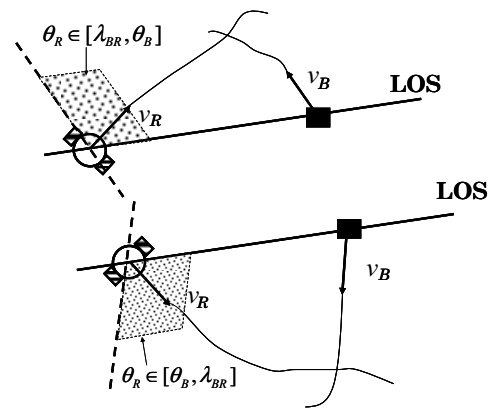


Fig. 3: Paths intersection illustration

Paths intersection does not imply collision. However, it is a necessary condition for collision. Next, we discuss the paths intersection by using the instantaneous values of the orientation angles of the robot and the moving object. We have the following result concerning the paths intersection:

Proposition 1: For $v_B>0$, $v_R>0$, the paths intersection can be characterized by the following conditions

$$C = \frac{\sin(\theta_R - \theta_B)}{\sin(\theta_R - \lambda_{BR})} < 0 \qquad (8)$$

and

$$D = \frac{\sin(\theta_R - \theta_B)}{\sin(\theta_B - \lambda_{BR})} < 0 \qquad (9)$$

From (8) and (9), the sign of $\sin(\theta_R-\lambda_{BR})$ and $\sin(\theta_B-\lambda_{BR})$ is the same when the paths intersect. Due to space limitation, the proof is not given here. However, interested reader is referred to [19].

Note that (8) and (9) are valid for constant and time varying orientation angles and line of sight angle. The robot and the moving object are in a collision course if they will arrive at the paths intersection point P at the same time. Now we discuss the collision course based on the relative kinematics equations. Our main result is stated as follows

Proposition 2: Let K be a constant. If (8) and (9) are satisfied for $v_R>0$, $v_B>0$, with

$$\frac{\dot{y}_d}{\dot{x}_d} = K \qquad (10)$$

then, the robot and the moving object are in a collision course; that is they will reach point P at the same time.

It can be shown by taking the time derivative in equation (4) that

$$K = \tan\lambda_{BR} \qquad (11)$$

is constant. Therefore equation (10) can be written as

$$\dot{y}_d = \dot{x}_d \tan\lambda_{BR} \qquad (12)$$

Also, because $\tan(\lambda_{BR})$ is constant, λ_{BR} is also constant; and therefore the time rate of λ_{BR} is zero. Equation (18) states that the relative distance between the robot and the moving object projected on the y-axis is proportional to the relative distance projected on the x-axis, with a constant proportionality factor. The proof for proposition 2 can be stated as follows:

Proof: The proof is based on the relative range r_{BR}. Transformation to polar coordinates of systems (1) and (2) yields the following system

$$\begin{aligned} \dot{r}_R &= v_R\cos(\theta_R - \lambda_R) \\ r_R\dot{\lambda}_R &= v_R\sin(\theta_R - \lambda_R) \\ \dot{\theta}_R &= \omega_R \end{aligned} \qquad (13)$$

for the robot, where (r_R,λ_R) are the robot's coordinates in polar coordinates, and

$$\begin{aligned} \dot{r}_B &= v_B\cos(\theta_B - \lambda_B) \\ r_B\dot{\lambda}_B &= v_B\sin(\theta_B - \lambda_B) \end{aligned} \qquad (14)$$

for the moving object, where (r_B,λ_B) are the moving object's coordinates in polar coordinates. Similarly to (6), the relative velocity can also be decomposed into two components along and across the line of sight LOS as follow

$$\vec{v}_{BR} = v_{BR}^{\perp}\vec{U}_{\perp} + v_{BR}^{\parallel}\vec{U}_{\parallel} \qquad (15)$$

where $U_\perp$ and $U_\parallel$ are the unit vectors across and along the line of sight, respectively. By considering systems (13) and (14), we get for the relative velocity components

$$\begin{aligned} \dot{r}_{BR} &= v_B\cos(\theta_B - \lambda_{BR}) - v_R\cos(\theta_R - \lambda_{BR}) \\ r_{BR}\dot{\lambda}_{BR} &= v_B\sin(\theta_B - \lambda_{BR}) - v_R\sin(\theta_R - \lambda_{BR}) \end{aligned} \qquad (16)$$

Recall that (8) and (9) result from the paths intersection conditions. From equation (10), since $\tan(\lambda_{BR})$= constant, it results that the rate of change of λ_{BR} is zero. This gives

$$v_R\sin(\theta_R - \lambda_{BR}) = v_B\sin(\theta_B - \lambda_{BR}) \qquad (17)$$

By using equation (17) and the equation for the relative velocity between the robot and the moving object, we get

$$\begin{aligned} \dot{r}_{RB} &= v_R\frac{\sin(\theta_R - \lambda_{BR})}{\sin(\theta_B - \lambda_{BR})}\cos(\theta_R - \lambda_{BR}) \\ &- v_R\cos(\theta_R - \lambda_{BR}) \end{aligned} \qquad (18)$$

By using geometric identities, we get in terms of the robot-moving object range.

$$\dot{r}_{BR} = v_R\frac{\sin(\theta_R - \theta_B)}{\sin(\theta_R - \lambda_{BR})} = Cv_R \qquad (19)$$

Or in terms of the velocity of the moving object v_B

$$\dot{r}_{BR} = v_B\frac{\sin(\theta_R - \theta_B)}{\sin(\theta_R - \lambda_{BR})} = Dv_B \qquad (20)$$

Since C and D are negative, the range rate is negative for forward motion. ■

When the paths intersection conditions with equation (10) are satisfied, the relative range r_{BR} changes sign at the collision point, i.e., when $r_{BR}=0$. Also, when this point is reached, the line of sight angle λ_{BR} switches by π (the new value of λ_{BR} equals the old value + π), and C and D become positive. Equations (8) and (9) give the paths intersection conditions for forward motions only. When both C and D are positive, the paths intersection takes place for backward motions (negative values for the linear velocities). The paths

intersection and the collision course equations are functions of the linear velocities and the orientation angles of R and B. Therefore, the collision course detection can be accomplished by using directional and speed sensors.

IV. COLLISION COURSE WHEN THE ROBOT AND THE MOVING OBJECT ARE MOVING IN THE SAME PATH

This particular case is shown in figure 4. In figure 4– a, R and B are both engaged in a pure pursuit towards each other. That is

$$\begin{aligned} \mathrm{LOS}\times\vec{v}_R = 0, \mathrm{LOS}\cdot\vec{v}_R > 0 \\ \mathrm{LOS}\times\vec{v}_R = 0, \mathrm{LOS}\cdot\vec{v}_R < 0 \end{aligned} \quad (20)$$

The robot and the moving object are in a collision course $\forall v_R$, $\forall v_B$. In figure 4-b, R is engaged in a pure pursuit and B is engaged in a pure escape. That is

$$\begin{aligned} LOS\times\vec{v}_R = 0, LOS\cdot\vec{v}_R > 0 \\ LOS\times\vec{v}_B = 0, LOS\cdot\vec{v}_B > 0 \end{aligned} \quad (21)$$

The robot and the moving object are in a collision course when $v_R > v_B$.

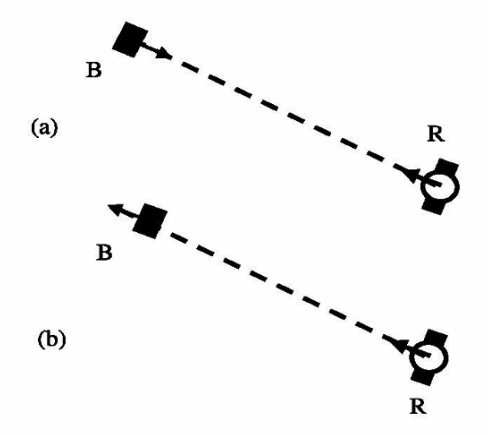


Fig 4. Collision course when the robot and the moving object move in the same path

The exact collision course is characterized by constant value for the of sight angle as described in equation (10). This leads us to the following important properties:

(a) When the exact collision course equation is satisfied, the motion of the moving object is a straight line as seen by the robot.
(b) The lines of sight are parallel to the initial line of sight. This is illustrated in figure 5.

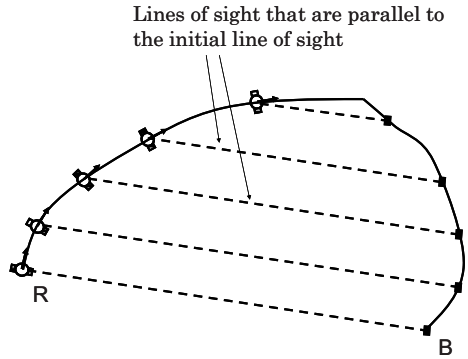


Fig. 5: An illustration of the exact collision course where the line of sight are parallel to the initial line of sight.

V. EXAMPLES OF COLLISION COURSE

We consider several examples illustrating the detection of the collision course between the robot and the moving object. Examples 1 and 2 illustrate the collision course, where the collision course equations are satisfied and example 3 illustrates a false alarm.

Example 1: R and B are moving in a straight line with constant speeds Consider the scenario of figure 6, with $\theta_B=0^0$, $\theta_R=45^0$,, $(x_B(t_0),y_B(t_0))=(0,5)$; $(x_R(t_0),y_R(t_0))=(20,0)$; v_B =1m/s, $\lambda_{BR}(t_0) = 166^0$.

—Verification of the paths intersection conditions: We have from equations (8) and (9) C = -0.8246, D = -2.9155

— Collision course: According to equation (25), R and B are in a collision course when v_R =0.2828m/s. Simulation of the motion of B and R (with v_R = 2.2828m/s) shows the collision course. The collision course can be deduced at the initial time.

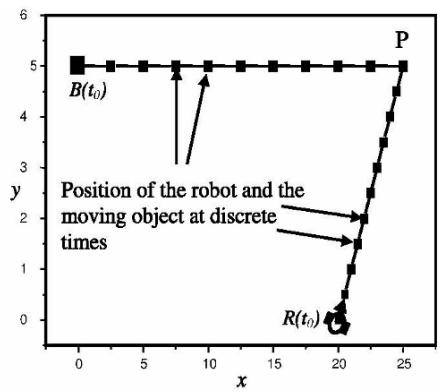


Fig. 6: Robot and moving object in a collision course, scenario of example 1.

Example 2: R is moving in a circle with constant speed and B in a straight line. This scenario is shown in figure 7, the robot moves with a constant speed v_R =5m/s. The initial positions $(x_B(t_0),y_B(t_0))$ =(0,0); $(x_R(t_0),y_R(t_0))$=(-10,25). The collision course takes place when the moving object moves according to the velocity profile shown in figure 8.

Example 3: False alarm
Consider again the scenario of figure 6, with v_R = 0.3328m/s. The range rate is negative in most of the time interval $[t_0, t_N]$. However, the collision does not take place as shown in the simulation. This can be predicted by using equation (10), which is not satisfied in this case. Therefore, this equation allows the detection of the false alarms.

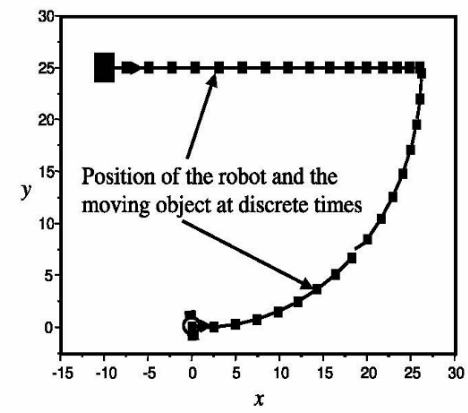


Fig. 7: Robot and moving object in a collision course, scenario of example 2

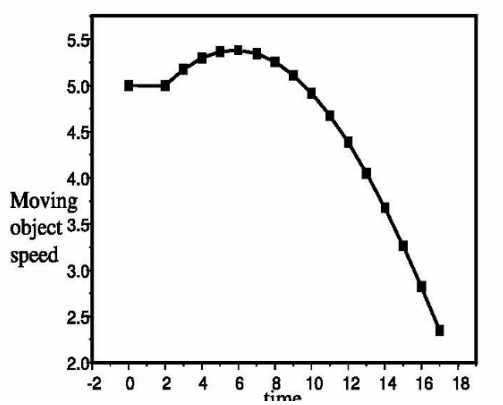


Fig 8: Velocity profile for the moving object under which the collision course conditions are satisfied.

VI. CONCLUSION

In this paper, we discussed the collision course conditions for a robot moving in the horizontal plane with a moving object. The collision course conditions are derived by using the relative kinematics equations. The collision course is expressed in terms of the linear velocities and the orientation angles of the robot and the moving object. It turns out that when the paths intersection conditions are satisfied, the collision course takes place when the projection of the relative distance robot-object on the y-axis is proportional to the projection of the relative distance on the x-axis, with a constant proportionality factor. Simulation examples confirm our theoretical results.

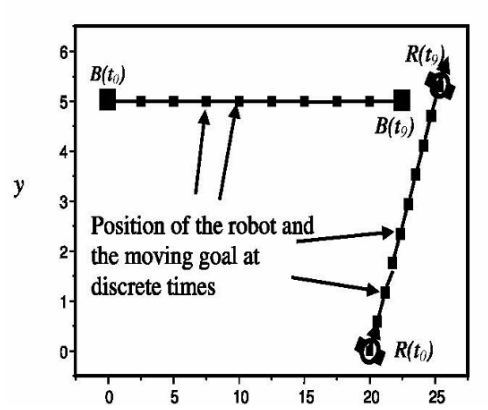


Fig. 9: Robot and moving object motions for the scenario of example 3.

ACKNOWLEDGMENT

The authors would like to thank Zinou Belkhouche for his inputs and suggestions.

REFERENCES

[1] O. Khatib, "Real-time obstacle avoidance for manipulators and mobile robots," International Journal of Robotics Research, vol. 5, no. 1, pp. 90–98, 1986.
[2] E. Rimon and D. Kopditschek, "Exact robot navigation using artificial potential functions," IEEE Transactions on Robotics and Automation, vol. 8, no. 5, pp. 501–517, 1992.
[3] J. Borenstein and Y. Koren, "Histogramic in-motion mapping for mobile robot obstacle avoidance," IEEE Transactions on Robotics and Automation, vol. 7, no. 4, pp. 535–539, 1991.
[4] J. Borenstein and Y. Koren, "The vector field histogram -fast obstacle avoidance for mobile robots," IEEE

Transactions on Robotics and Automation, vol. 7, no. 3, pp. 278–288, 1991.
[5] J. Latombe, Robot motion planning. Massachusetts: Kluwer, 1991.
[6] S. Ge and Y. Cui, "Dynamic motion planning for mobile robots using potential field method," Autonomous Robots, vol. 13, pp. 207–222, 2002.
[7] T.Fraichard and C. Laugier, Dynamic trajectory planning, path-velocity decomposition and adjacent paths in Proc. IEEE International Conference on Robotics and Automation, Atlanta, GA, May 1993, pp. 40–45.
[8] T. Fraichard, "Trajectory planning amidst moving obstacles: path-velocity decomposition revisited," Journal of the Brazilian Computer Science Society, special issue on Robotics, vol. 4, no. 3, pp. 5–13, 1998.
[9] H. Kao-Shing and J. Ming-Yi, "Automatic generation of a collision free speed profile for the maneuvering motion," in Proc. IEEE International Conference on Systems, Man, and Cybernetics, Tokyo, Oct. 1999, pp. 708–713.
[10] G. DeSouza and A. Kak, "Vision for mobile robot navigation: A survey," IEEE Transactions on Pattern analysis and machine intelligence, vol. 24, no. 2, pp. 237–267, 2002.
[11] A.Fujimoriand S. Tani, "A navigation method of mobile robots with collision avoidance for moving obstacles," in Proc. IEEE International Conference on Industrial Technology, Bangkok, Thailand, Dec. 2002, pp. 1–6.
[12] A. Fujimori, M. Teramoto, P. Nikiforuk, and M. M. Gupta, "Cooperative collision avoidance between multiple mobile robots," Journal of Robotic Systems, vol. 17, no. 7, pp. 347–363, 2000.
[13] C. Wu, K. F. D. Lee, and K. Hwang, "An automated collision-avoidance motion planner among moving objects or machines," in Proc. IEEE Conference on Decision and Control, Brighton, Dec. 1991, pp. 2422–2427.
[14] H. Ma, D. Cannon, and S. Kumara, "A scheme integrating neural networks for real-time robotic collision detection," in Proc. IEEE International Conference on Robotics and Automation, Nagoya, May 1995, pp. 881–886.
[15] T. Masumoto and K. Kosuge, "Dynamic collision detection method using adaptive control law,"in Proc. IEEE International Conference on Industrial Electronics, Nagoya, Oct. 2000, pp. 2243–2248.
[16] K. Valavanis, T. Hebert, R. Kolluru, and N. Tsourveloudis, "Mobile robot navigation in 2-d dynamic environments using an electrostatic potential field," IEEE Transactions on Systems, Man and Cybernetics, Part A, vol.30, no.2,pp. 187–197, 2000.
[17] A. del Pobil, M. Perez, and B. Martinez, "A practical approach to collision detection between general objects," in Proc. IEEE International Conference on Robotics and Automation, Minneapolis, MN, Apr. 1996, pp. 779–784.
[18] E. Bernabeu and J. Tornero, "Hough transform for distance computation and collision avoidance," IEEE Transactions on Robotics and Automation,, vol. 18, no. 3, pp. 393–398, 2002.
[19] F. Belkhouche and B. Belkhouche, Kinematics-based characterization of the collision course, to appear in the International Journal of Robotics and Automation.

Acquisition System for Monitoring Vibrations

Grofu Florin "Constantin Brancusi" University – Tg-Jiu
Popescu Luminita "Constantin Brancusi" University – Tg-Jiu
Cercel Constantin "Constantin Brancusi" University – Tg-Jiu

Abstract: **Antifriction bearing failure is a major factor in failure of rotating machinery. As a fatal defect is detected, it is common to shut down the machinery as soon as possible to avoid catastrophic damages. Performing such an action, which usually occurs at inconvenient times, typically results in substantial time and economical losses. It is, therefore, important to monitor the condition of antifriction bearings and to know the details of severity of defects before they cause serious catastrophic consequences. The vibration monitoring technique is suitable to analyze various defects in bearing. This technique can provide early information about progressing malfunctions. In this paper is presented a data acquisition system conceived for the analysis of the signal from two vibration transductor.**

Keywords: Antifriction bearings; Prediction; Vibration signal; Acquisition System

I. Introduction

Condition monitoring of antifriction bearings in rotating machinery using vibration analysis is a very well established method. It offers the advantages of reducing down time and improving maintenance efficiency. The machine need not be stopped for diagnosis. Even new or geometrically perfect bearings may generate vibration due to contact forces, which exist between the various components of bearings. Antifriction bearing defects may be categorized as localized and distributed. The localized defects include cracks, pits and spalls caused by fatigue on rolling surfaces. The other category, distributed defects include surface roughness, waviness, misaligned races and off size rolling elements. These defects may result from manufacturing error and abrasive wear. Antifriction bearing failures result in serious problems, mainly in places where machines are rotating at constant and high speeds. To prevent any catastrophic consequences caused by a bearing failure, bearing condition monitoring techniques, such as, vibration analysis and acoustic emission analysis have been developed to identify existence of flaws in running bearings.

Vibration signature monitoring and analysis in one of the main techniques used to predict and diagnose various defects in antifriction bearings. Vibration signature analysis provides early information about progressing malfunctions and forms the basic reference signature or base line signature for future monitoring purpose. Defective rolling elements in antifriction bearings generate vibration frequencies at rotational speed of each bearing component and rotational frequencies are related to the motion of rolling elements, cage and races. Initiation and progression of flaws on antifriction bearing generate specific and predictable characteristic of vibration. Components flaws (inner race, outer race and rolling elements) generate a specific defect frequencies calculated from equations, mentioned by Chaudhary and Tandon, namely:

Inner face malfunction frequency

$$f_i = \frac{n}{2} f_r \left[1 + \left(\frac{BD}{PD} \right) \cos \beta \right] \quad (1)$$

Outer race malfunction frequency

$$f_o = \frac{n}{2} f_r \left[1 - \left(\frac{BD}{PD} \right) \cos \beta \right] \quad (2)$$

Roller malfunction frequency

$$f_R = \frac{BD}{PD} f_r \left[1 - \left(\frac{BD}{PD} \right)^2 \cos^2 \beta \right] \quad (3)$$

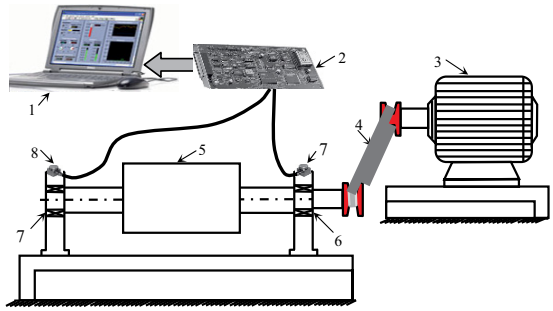
where: BD roller diameter, PD pitch diameter, *fr* rotational frequency, *n* number of rollers and ß angle of contact.

The time domain and frequency domain analyses are widely accepted for detecting malfunctions in bearings. The frequency domain spectrum is more useful since it also identifies the exact nature of defect in the bearings.

II. Experimental Stand

An experimental test rig built to predict defects in antifriction bearings is shown in Figure 1. The test rig consists of a shaft with central rotor, which is supported on two bearings. An induction motor coupled by a flexible coupling drives the shaft. Self aligning double row ball bearing is mounted at driver end and cylindrical roller bearing is mounted at free end. The cylindrical roller bearing is tested at constant speed of 1400 rpm. Cylindrical roller bearing type 6308C3 (with outer race and roller defects and with inner race defect) has been used for analysis.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 89–92.



1-PC; 2- Data Acquisition Sistem; 3-Motor; 4–Flexible Coupling;
5–Rotor; 6–Self Aligning Ball Bearing; 7–Roller Bearing; 8–Transductors

Fig.1

The details of the bearings used in the present analysis are : number of roller 8, outer diameter 90mm, inner diameter 40mm, pitch diameter 55mm, roller diameter 15mm, contact angle $\beta=0°$

In order to make the acquisition of the useful signal, a vibration transducer was put on each bearing. The signal from the two sensors is acquisitioned and transmitted towards a computer which will do the frequency analysis of the received signals. The block scheme of the acquisition part is presented in the following figure (fig.2).

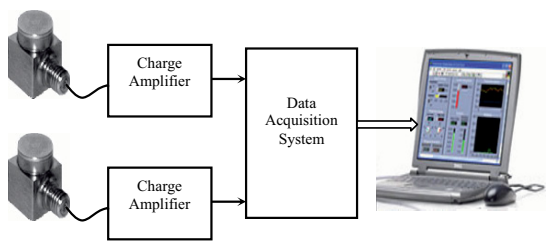


Fig. 2

III. Vibration Transducer

The vibration transducer that is used is MAQ 36. Model MAQ36 charge output miniature accelerometer is designed to be used in Industrial test and automation environments; including laboratory testing, modal studies and test cells where high temperatures are likely to be encountered and where space is limited and small size is desired or a high natural frequency is required. The MAQ36 is a self generating piezoelectric transducer which has no internal electronics and requires no external power for operation. These units are usually connected to a local charge amplifier that is mounted as near as possible in a lower temperature environment. The seismic element is mechanically isolated from the mounting base, resulting in low base strain sensitivity. The stainless steel materials are non-magnetic resulting in very low magnetic field susceptibility. These features, together with a sealed body, assure accurate and reliable data.

The accelerometer is the transducer type most commonly encountered when measuring vibration levels. The accelerometer normally consists of a seismic mass mechanically connected to the accelerometer base through a piezoelectric material. Piezoelectric materials have the property of producing electrical charge when bent and twisted (even shear forces will work here). Basically, a charge is generated. Hence, a charge sensitive amplifier will produce a signal independent of the capacitance of the cable between the accelerometer and the amplifier.

IV. Acquisition System

To analyze the vibrations, the frequency analysis method is used, which implies the implementation on a PC of the Fourier transform. In order to have the most correct analysis, the acquisition rate of the useful signal must be as big as it can report to the maximum frequency of the analyzed signal's spectrum. Making such an acquisition system implies to fulfill simultaneously more conditions which most of the times are in contradiction.

First, a very high speed of acquisition is required, which implies a big volume of data. To make a real time analysis of these data, there must be used complex systems with high power of calculus, like the process computers. The vibrations produced in the bearings which must be analyzed, appear in an improper industrial environment. Regarding the climate and mechanical emplacement, the usual process computers cannot be positioned next to the process.

The alternative would be placing an acquisition system next to the two sensors and transmitting the information in a numerical way to the computer, which is placed in the control room. In this case, a high speed numerical transmission is necessary, so that the analysis of the acquisitioned signals to be in real time. Generally, data transmission at high speeds can be made through parallel interfaces (with a big number of wires) or through core complicated serial interfaces.

Because the defects that appear in the bearings is a slow phenomenon, for the analysis of the defects using vibrations analysis, a fast signal acquisition can be made for a certain period of time with local hold of the data and then a slower transmission of the data through a standard serial interface (for example RS 485). The acquisition system must be able to make acquisitions of at least 100.000 samples/second and to hold a big number of samples (the equivalent of an interval of few seconds to observe the useful signal). The structure of the proposed acquisition system is presented in figure 3:

IV.1. Core with Microcontroller

The core with microcontroller is made with DS89C420. The DS89C420 offers the highest performance available in 8051-compatible microcontrollers. It features a redesigned processor core that executes every 8051 instruction (depending on the instruction type) up to 12 times faster than the original for the same crystal speed. Typical applications see a speed improvement of 10 times using the same code and crystal. The DS89C420 offers a maximum crystal speed of

33MHz, achieving execution rates up to 33 million instructions per second (MIPS).

DS89C420 is 8051 Pin- and Instruction-Set Compatible and have four Bidirectional I/O Ports, Three 16-Bit Timer Counters, 256 Bytes Scratchpad RAM, 16kB Flash Memory, In-System Programmable through Serial Port, Dynamically Adjustable by Software, 1 clock-per-machine cycle and Single-cycle instruction in 30ns, Optional variable length MOVX to access fast/slow peripherals, Dual data pointers with auto increment/decrement and toggle select, Programmable clock divider, Two full-duplex serial ports, Programmable watchdog timer, 13 interrupt sources (six external), Five levels of interrupt priority.

IV.2. RAM Memory

The RAM memory used for the local hold of the conversions result is of the DS1270Y type.

The DS1270Y 16Mb Nonvolatile SRAMs are 16,777,216 bit, fully static nonvolatile SRAMs organized as 2,097,152 words by 8 bits and read and write access times as fast as 70 ns. Each NV SRAM has a self-contained lithium energy source and control circuitry which constantly monitors VCC for an out-of-tolerance condition. When such a condition occurs, the lithium energy sources is automatically switched on and write protection is unconditionally enabled to prevent data corruption. There is no limit on the number of write cycles which can be executed and no additional support circuitry is required for microprocessor interfacing. Optional industrial temperature range of -40°C to +85°C, designated IND.

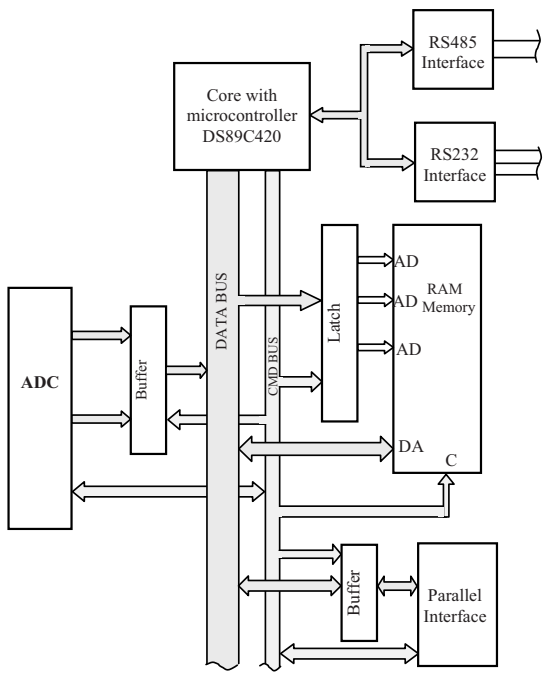


Fig. 3

IV.3. Analog-to-Digital Converter

The analog-to-digital converter which is used is MAX 120. The MAX 120 is BiCMOS, sampling 12-bit analog-to-digital converters (ADCs) combine an on-chip track/hold (T/H) and low-drift voltage reference with fast conversion speeds and low power consumption. The T/H's 350ns acquisition time combined with the 1.6µs conversion time results in throughput rates as high as 500k samples per second (ksps). The MAX 120 accepts analog input voltages from -5V to +5V and operates with clocks in the 0.1MHz to 8MHz frequency range.

The MAX 120 employ a standard microprocessor (µP) interface. Three-state data outputs are configured to operate with 12-bit data buses. Data-access and bus-release timing specifications are compatible with most popular µPs without resorting to wait states. In addition, the MAX 120 can interface directly to a first in, first out (FIFO) buffer, virtually eliminating µP interrupt overhead.

The data transfer between the analog-to-digital converter and memory or between the memory and the parallel interface can be made through the microcontroller or directly on the data bus. In this case, the microcontroller generates only command signals which are necessary for data transfer. To increase the acquisition speed, saving previous data is done simultaneously with a new conversion. In this way, speeds of 400.000 samples/second can be reached. Transmitting data from the RAM memory towards the PC can be done serial (RS 485) for big distances, or serial (RS232) or parallel on small distances.

V. Software

The acquisition program is made with the help of LabWindows/CVI 7.1 and is split into 3 components:

- one component which makes the acquisition and data process and which is conceived as a "service" under Windows XP;
- one MySQL component (data base server) which asynchronously takes over the files stored on the HDD by the first component and stores them into a data base;
- one component which is launched at demand and which presents the data in the data base into a suggestive graphical interface.

If the acquisition board is placed at a considerable distance from the PC, the acquisition is made through serial RS232 by a RS485/RS232 converter. The program works on interrupt, meaning to receive data in the RS 232 serial buffer, and when it is full (or we can set a certain number of characters, smaller than the buffer's dimension), an interrupt is generated. The function which deals with the interruption downloads the buffer of the serial into a memory variable of string type as in the example: unit data[100.000], where data[i] = [0x0F& high_octet] *256 + low_octet meaning high octet has the first n bits equal to zero (the analog-to-digital converter works on 12 bits).

These data from the memory are processed in the following way:

- applying the Fourier transform, will result the frequency spectrum of the data from the memory.
- from the frequency spectrum analysis, a diagnosis of the device try is made, meaning the type of defect (for example, for the bearings: broken ball, deficient oiling, usage, the engine and pump are not on the center).
- the conclusions from the previous step are being transmitted to the dispatcher for a real time information on the defects. If the found defects are major, the "emergency stop" of the device is necessary.
- the frequency characteristic from step 1 is memorized in the computer's memory.

According to the diagnosis algorithms, there may be certain frequency domains which must be analyzed normally, thing that can be done with performant digital filters. If defects are found, the entire acquisitioned frequency spectrum is analyzed. This is done not to have high time values for processing or to eliminate parasite signals not related to the process.

These times for acquisition, processing, memorizing, must be included into the sampling period chosen for the process if the same computer coordinates the whole process. With the help of the memorized files, a history of the functioning can be made, for two purposes:

- visualizing in time the evolution of the device, useful for revision and repairs planning.
- creating and improving some algorithms for defects identification corresponding to the analyzed physic device.

CONCLUSION

Time waveform and frequency spectrum provide useful information to analyze defects in antifriction bearings. Time waveform indicates severity of vibration in defective bearings.

Frequency domain spectrum identifies amplitudes corresponding to defect frequencies and enables to predict presence of defects on inner race, outer race and rollers of antifriction bearings. The distinct and different behavior of vibration signals from bearings with inner race defect, outer race defect and roller defect helps in identifying the defects in roller bearings.

Also, the acquisition system presented ensures the obtaining of some precise acquisitions which correspond from the point of view of resolution (number of bits) and of speed of acquisition. It's simple and robust structure allows this system to be made according to any industrial environment.

After the numerical processing of the data through filtering and fast Fourier transform is optined the power specters for experimental model corresponding to a good ball bearing and to a damaged one.

From the analysis of the two power specters, it results that we have a component given by the rotation frequency of the axel (f_0=28 Hz), which has approximately the same value for the two ball bearings. We observe that for the good bearing there are other spectral components, but smaller, and for the damaged bearing there are spectral components even higher than the fundamental component, f_0. These spectral components appear at lower than f_0 frequencies and at higher ones.

Analyzing power specters for the damaged ball bearing we observe the existence of the components with very close frequencies to the ones calculated theoretically, but also of other spectral components. Only on the basis of the theoretical analysis of the observed data, the following conclusion can be drawn about the damaged ball bearing: has its cage very deteriorated (visible with the naked eye), has the exterior trajectory and one of the balls with faults and a powerful off centering and mechanical clearances in the gearing.

An exact diagnosis can be made through the practical study of a big number of ball bearings with only one fault, study of the resultant spectral components and creating a data base corresponding to that type of gearing and ball bearing.

REFERENCES

[1]. Y Li and C Zhang. 'Dynamic Prognostic Prediction of Defect Propagation on Rolling Element Bearing'. *Journal of Vibration and Acoustics*, *Trans of ASME*, vol. 120, no 1, pp 214-220.

[2]. Teruo Igarishi and Hiroyoshi. 'Studies on Vibration and Sound of Defective Rolling Bearings'. *Bulletin JSME*, vol. 25, no 204, 1980, pp 994-1001.

[3]. I J Taylor. 'Identification of Bearing Defects by Spectral Analysis'. *Journal of Mechanical Design*, *Transaction of ASME*, vol. 120, 1980, pp 199-204.

[4]. Porat B., *A Course in Digital Signal Processings*, John Wiley & Sons, Inc., New York, 1997.

[5]. D. H. Sheingold, *Analog-Digital Conversion Handbook*, Analog Devices, Inc.

[6]. J Jacob Wikner, *STUDIES ON CMOS DIGITAL-TO-ANALOG CONVERTERS,* Linköping Studies in Science and Technology Dissertation No. 667, Sweden Linköping 2001

[7]. Maxim, *Digital-Analog Converters Are a "Bit" Analog,* Application Note 1055: Apr 16, 2002

[8]. Nicolae Tomita, *Întreținerea preventivă și predictivă*, http://www.vibrotest.com/ 2005

[9]. REBAM**,** *Predictive Maintenance through the Monitoring and Diagnostics of Rolling Element Bearings***,** REBAM Application Note 2005

[10]. M Amarnath, R Shrinidhi, A Ramachandra, S B Kandagal, *Prediction of Defects in Antifriction Bearings using Vibration Signal Analysis,* IE(I) Technical Journals : Mechanical Engineering -July 2004

[11]. Honeywell Sensotec, *Accelerometers Frequently Asked Questions*, www.sensotec.com, 2003

[12]. Thomas B. Cho, *Circuit Techniques for Low Power, High Speed Pipelined A/D chapter 2,* University of California, Berkeley 2000

[13]. Dallas Semiconductor, *Ultra-High-Speed Flash Microcontroller User's Guide* , http://www.maxim-ic.com/ 2004

[14]. Maxim-IC, *App Note 604: Fast Memory Transfers with the Ultra High-Speed Microcontroller - DS89C420,*

Object-of-Interest Selection for Model-Based 3D Pose Tracking with Background Clutter

Hans de Ruiter, *Member, IEEE*, and Beno Benhabib

Abstract— **Extensive research on model-based 3D object tracking has yielded a number of effective methodologies. However, work reported in the literature on initiating tracking has been limited. This paper addresses this issue via a novel framework that can automatically find an *Object-of-Interest* (OI) in a dynamic scene and initiate tracking. Since OI definition is, typically, application dependent, the proposed framework is modular and customizable. It combines a real-time motion segmentor with a set of customizable interest filters to separate, highlight, and select the OIs.**

Furthermore, our earlier model-based object tracker is extended in this paper to utilize OI-selection data and track objects in the presence of background clutter. Thus, the overall computer-vision system presented in this paper can automatically select, model, and track the six degree-of-freedom position and orientation of an OI, whose model is not known *a-priori*.

Proposed algorithms were verified via extensive simulations and experiments, some of which are presented herein.

I. Introduction

There has been extensive research reported on real-time, model-based, 6-dof (degree-of-freedom) object tracking (e.g.,[1]-[10]). These object trackers have been proposed for use in robotic-control systems, allowing greater interactivity between the (mobile) robot and its environment. However, although such fully autonomous systems should be able to decide which object(s) they should track, automatically, the subject of tracking initialization has often been ignored or avoided.

This paper addresses selection of an Object of Interest (OI) within a scene containing multiple objects and background clutter. The proposed computer-vision system automatically selects an OI in real-time and interfaces seamlessly with the latest version of our object tracker, as described in this paper (whose earlier versions were detailed in [1]-[3]).

The OI-selection algorithm bears some similarity to visual-attention algorithms, such as presented by Sun and Fisher [11] and Ahrns and Neumann [12]. Both OI selection and visual attention algorithms try to highlight interesting parts of an image in order to reduce calculations downstream. There are, however, two key differences: an OI selector has to select an object to track/analyse, not just highlight interesting regions as in a visual-attention algorithm; and, our goal is to select objects that are of interest specifically to the system (i.e., to the end application), instead of searching with more general criteria.

A key difficulty with OI selection is that what constitutes an OI is application dependent. For some systems, an OI may be a human subject for identification via face recognition; for others, an object moving toward the robot may be of greatest interest; for others still, the system may be looking for a particular type of object, or an object within a particular region (such as a part on a production line). In order to cope with this variability, the proposed OI-selection system uses a novel modular set of customizable Interest Filters (IFs).

A secondary but important further contribution of this paper is the extension of our earlier object-modeller algorithm to multi-object environments with cluttered backgrounds. Earlier in [3], a 3D modeller was used to enable 6-dof pose tracking of *a priori* unknown objects. However, the modeller could only cope with a single object being visible in front of a uniform background. In this paper, the IF framework, utilized by the OI selector, is used to highlight regions of the image belonging to the tracked OI, thus, enabling the modeller to separate the OI from its surroundings.

Combining the OI selector with the improved modeller and object tracker yields an effective real-time tracking system that can autonomously select an OI and proceed to track it. The proposed system has been tested on synthetic and real-world motion sequences.

II. Object Tracker – An Overview

The proposed methodology uses a novel 3D model-based approach in order to track the full 6-dof pose of an object. It uses a real-time modeller to facilitate the tracking of objects whose models are not known *a priori*. The object's pose (position and orientation) is estimated by projecting a visual 3D model of the *target* object onto the image plane for comparison with the input data. Motion between the object's predicted pose and its actual pose is calculated using optical-flow between the projected image and the real target object's image. This motion is, then, used to correct the predicted pose, yielding the final pose estimate.

3D graphics hardware via OpenGL [13] is used in our work to perform projection and other image-processing

Manuscript received October 14, 2007. This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Hans de Ruiter and Beno Benhabib are with the University of Toronto, Toronto, Ontario, Canada (e-mail: deruiter@mie.utoronto.ca, beno@mie.utoronto.ca).

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 93–98.

operations, relieving the main processor from this task and, hence, significantly improving the performance of the system. The algorithm also exploits the depth-map and mask produced by OpenGL for motion calculation and segmentation, respectively.

The modeller rapidly builds a model of the visible part of an object. It uses a real-time depth-map extraction algorithm to extract 3D geometry and projective texture-mapping [14] to represent surface features. The model is rebuilt periodically in order to cope with object rotation.

A full explanation of the tracking algorithm can be found in our previous papers [1]-[3]. This paper improves this modeller, allowing the object tracker to track objects in front of a cluttered (i.e., non-uniform) background.

III. Proposed Methodology – Object-of-Interest Selection and Modelling

The proposed OI-selection methodology is based on a cascading framework of customizable Interest Filters (IFs). The IF framework outputs a map of regions that are *of interest*. This information, coupled with a segmentation map generated by a motion-segmentation algorithm, is used to identify which region is *of greatest interest* to the overall autonomous system. The OI region obtained is passed on to the object modeller for modelling and tracking. In essence, the OI selector is the object-tracking system's start-up module.

A. Interest-Filter Framework

The core of the OI-selection system is the IF framework, which serially connects individual IFs into a filter bank. Each IF filter's input data is based on a set of criteria (such as colour, motion, etc.). IFs also take the previous filter's interest map as input and reduce the total area of interest. The first filter in the system receives a blank interest map filled with ones. An interest value of one signifies full interest; a value of zero signifies no interest at all. The end result is a set of highlighted interest regions.

An example IF bank is shown in Figure 1. The first filter reduces the region of interest based on colour; pixels are compared to a list of colours known to appear on the OI. The next two IFs filter based on position and velocity, respectively. Each IF requires different input data. For example, the colour IF operates on input images only, whereas the position IF requires a depth map as well.

Although each IF operates on data differently, there are a few rules that they all must conform to: an IF is not allowed to increase the interest value for a pixel, the interest value may either remain constant or decrease. A typical transformation would be:

$$S_n = S_{n-1} \cdot S_{n,calc} \,, \qquad (1)$$

where S_n is the interest value output by filter n and $S_{n,calc}$ is the interest value calculated by the filter operation for filter n. An IF can be designed simply by developing an

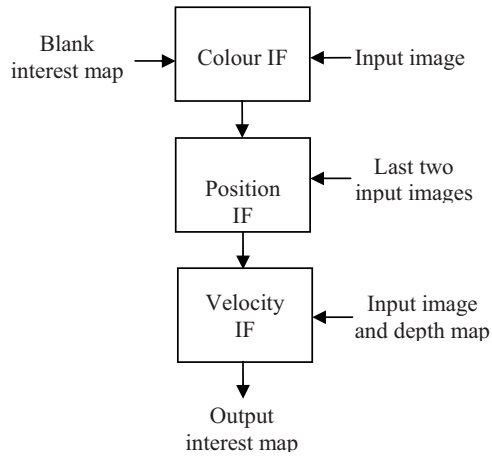


Figure 1: An IF bank with multiple, serially connected IFs.

algorithm that outputs the local interest value, $S_{n,calc}$; its interaction with other IFs in the filter bank is set by (1). As a general rule, IFs should be connected in decreasing order of processing speed (or increasing order of computational cost). They need to operate only on pixels whose input interest value (from the previous filter) is non-zero. This optimizes the overall speed of the filter bank as a whole.

Tailoring the IF framework to an application can be better explained via an example: Let us consider a room that Automated Guided Vehicles (AGVs) have free access to, but human entry is restricted only to those with a security clearance, Figure 2. An IF framework selects the OI as follows: a colour-based IF filters out floor and AGVs based on their colour, resulting in Figure 3(a); a positional IF focuses attention on objects close to the restricted-room entrance, Figure 3(b); a velocity-based IF highlights regions moving toward the entrance, Figure 3(c). Finally, the OI selector selects the region of greatest interest and, for example, a face-recognition system determines whether or not the individual has access rights, Figure 3(d).

B. Motion Segmentation

Due to interest regions being sparsely covered, interest maps are unsuitable for separating different objects. Some other

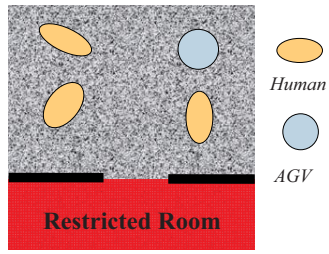


Figure 2: An example scenario requiring appropriate OI selection

method is required for segmenting the image into different regions (corresponding to objects). Image-segmentation

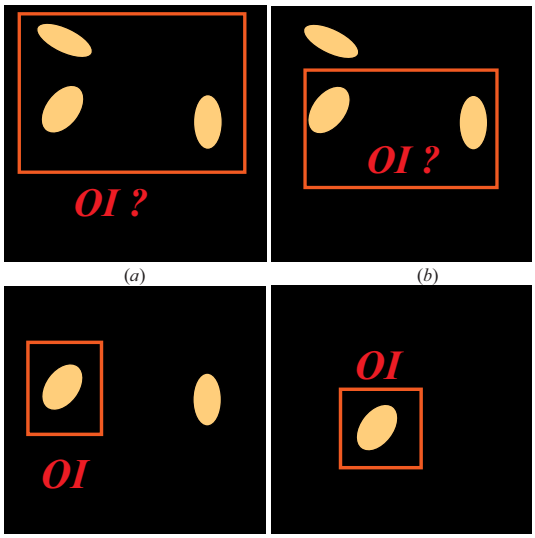


Figure 3: (*a*) A colour IF highlights humans; (*b*) a positional IF highlights regions close to the door; (*c*) a velocity IF highlights regions moving toward the restricted room, thus, (*d*) isolating the OI to track.

algorithms tend to be non-real-time and segment into regions of uniform colour. Objects, typically, are not of single colour, rendering image segmentation unsuitable for the task.

The proposed algorithm uses motion segmentation to divide the image up into separate objects. Regions within an image moving under different transformations generally belong to different objects (assuming that all objects are rigid). Thus, every region (or segment) can be treated as a separate object, or potential OI. Motion segmentation is used, because it is assumed that objects that are of interest will be moving. If one wishes to search for static objects, an alternative segmentation algorithm could be used within the overall OI selection framework presented herein. Due to the object being static, real-time operation would no longer be required. However, for the purpose of this paper, it is assumed that OIs are in motion and real-time selection is required.

The main requirement for the motion segmentor is that it operates in real-time. This is particularly important as several other operations also need to be performed simultaneously (e.g., interest filtering, modelling, and tracking). Approximate or even partial separation of an OI is acceptable. The current implementation utilizes 2D motion segmentation (Section IV.B). Although a 3D motion segmentor could provide more accurate segmentation, such segmentors are non-real-time at present.

C. OI Selection

Once interest and segmentation maps are available, an OI can be selected. The selection algorithm scans through the segmentation map, interest map, and depth map (including it's texturedness image, T), constructing a list of interest properties for each region. For every region (or segment) R_i, where $i \in [1..N]$, and N is the number of regions:

- $TotalArea_i = \sum_{x,y \in R_i} 1,$ (2)
- $InterestArea_i = \sum_{x,y \in V_i} 1,$ (3)
- $TotalInterest_i = \sum_{x,y \in V_i} S(x,y),$ (4)
- $MeanInterest_i = \frac{TotalInterest}{TotalArea},$ (5)
- $DepthPixelArea_i = \sum_{x,y \in W_i} 1,$ and (6)
- $MeanDepth_i = \sum_{x,y \in W_i} D(x,y).$ (7)

Above $S(x,y)$ and $D(x,y)$ are the interest value and depth (i.e., distance to the surface along the camera's optical axis, z) at (x,y), respectively. V_i is a sub-region to R_i in which $S(x,y) > \tau_s$, i.e., the interest level is greater than a threshold, and W_i is a sub-region of V_i containing all pixels belonging to V_i for which the texturedness is greater than a threshold, i.e., $T(x,y) > \tau_T$ (τ_T is set to the same value as used by the depth-map extractor). This additional restriction for pixels used to calculate the mean depth ensures that only pixels with reliable depth-values are used.[1] Additionally, the region's bounding-box is calculated. This is combined with the mean-depth, to estimate the region's physical width and height (as viewed by the camera).

With all the above properties calculated, the OI selector must now decide: whether an OI exists and, if so, which OI is of greatest interest. An OI exists if the following conditions are met: its physical width and height are within a predetermined range; its mean and total interest levels are above their respective thresholds; its total area is greater than a threshold; and, $V_i / R_i > \tau_A$ (i.e., the proportion of the OI region covered by interest pixels is high enough). The selected OI is the region R_i that satisfies the above conditions and has the highest mean interest level. This region is labelled R.

D. Initial-Pose Generator

Theoretically, an OI's reference frame can be placed anywhere relative to its surface. However, the optimal reference frame location can be application dependent. For example, in some applications, the geometric centre may be most appropriate; for others, the centre of mass, or even

some other point within the object. When tracking an object whose model is not known *a priori*, it could be difficult to find such an optimal reference frame. Nevertheless, a reference frame must be selected based on the available data.

The specific initial-pose generator algorithm used is not critical to operation of the object tracker. However, the reference frame should preferably be placed within the object, or close to the visible surface. The specific algorithm implemented in our laboratory is detailed in Section IV.C.

E. The New and Improved Modeller

With all the above steps completed, the selected OI region, generated initial pose (i.e., reference frame for the OI), interest-map, and the depth-map are passed to the object tracker. The modeller must now use this information to construct a model of the OI and pass it to the object tracker.

The modeller and object tracking system presented in [3] had to be modified herein before it could use the OI selection system. In particular, our earlier modeller did not cope with background clutter or a multi-object environment. Separating the OI from its environment is a difficult proposition. However, our IF framework has already highlighted the OI. Thus, our technique masks the texturedness image by the thresholded interest-map. Any pixel in the texturedness image for which the interest level is below τ_i, is set to zero. Other than using the masked texturedness image instead of the raw texturedness image, the modeller operates as per the original algorithm in [3]. Vertices selected for building the tessellated object model, only derive from points on the OI. When building the initial model, the texturedness image is also masked by region R; when rebuilding the model *n* frames later, modelling is restricted to the region covered by the old model's bounding box. Therefore, the model built covers the OI and not the entire image. The IF framework is used in both acquiring the initial model and in rebuilding it.

IV. Implementation

A. Implemented Interest Filters

Two IFs were implemented: a colour IF and an IF that performs background subtraction. The colour IF separates image regions that have a particular colour from the background. Its interest value is based on the following formula:

$$S_{n,calc} = clamp\left(\sum_{i=0}^{M} clamp\left(1 - \frac{\|\boldsymbol{I} - \boldsymbol{I}_i\|^2}{I_{i,md}^2}, 0, 1\right), 0, 1\right), \quad (8)$$

where $\boldsymbol{I}$ is the RGB colour value for the current pixel, $\boldsymbol{I}_i$ is the reference colour *i* to which it is being compared, and $I_{i,md}$ is the maximum Euclidean distance allowed between the pixel's colour and the reference colour *i*. There are *M* reference colours to which the input image is compared. The function *clamp* clamps values to the range 0 to 1, i.e.

$$clamp(x,a,b) = \begin{cases} x, x \in [a,b] \\ a, x < a \\ b, x > b \end{cases}, \quad (9)$$

with *a* and *b* set to 0 and 1, respectively.

The result of (8) is an interest value that is high when a pixel's colour is close to one of the desired colour values and low elsewhere. $I_{i,md}$ limits the radius around a desired colour that produces an interest value above zero.

Background subtraction is a simple and fast method of detecting moving objects. It operates by subtracting a reference image of the background from the current input image. All differences above a certain threshold are marked as moving. The background subtraction algorithm in this IF is based on the adaptive algorithm by Collins and Dennis [15] minus the region-filling algorithm used to fill in solid colour regions. Region filling was not required for our IF. Moving pixels were given an interest value of one; static pixels were given a value of zero.

B. The Motion Segmentor

The motion segmentor used is based upon the algorithm developed by Chung et al. [16]. It is block based and operates in real-time. The image is divided into 8×8 blocks, for which optical-flow is calculated. Adjacent blocks with similar motions are grouped into regions. Due to space restrictions, the details of the motion segmentor are not given here. However, one can note that the internal operation of the motion segmentor is not important to the understanding of the OI selection system. In practise, any motion segmentor could be used, with results varying based on the motion segmentor's effectiveness.

C. The Initial Pose Generator

A fast reference-frame generator has been developed. The reference-frame's orientation is aligned with the world reference frame. This was deemed as acceptable since the reference frame's orientation does not affect tracking. It is the responsibility of an external system to characterize the object and determine how to interact with it (e.g., pick a reference frame aligned with the object's shape).

The reference frame's position is generated as follows:

- The 2D mid-point of the region R 's bounding box is calculated, and
- The reference frame's position is set by projecting the 2D region mid-point back into the scene by the mean depth plus an offset.

This procedure places the reference frame about the middle of the visible object, just behind the 3D surface that

[1] Depth can only be measured in image regions with non-zero image gradients (i.e., textured regions).

will be generated by the modeller.

V. Experiments

Operation of the OI selector was tested using a series of motion sequences. In all experiments, the two implemented IFs were used. The colour IF was adjusted to match the OI. The OI Selector was tested within an overall object-tracking system, including the OI selector, modeller, and tracker modules. This tests the OI selector within its target application. Some of the experiments are given below. Five cameras were used for depth-map extraction.

In order to verify that the OI selector can operate in a multi-object environment, experiments were performed on a two object sequence (Figure 4). The colour IF was tuned first to the cube (Figure 5) and, then, to the globe (Figure 6). In each case, the OI selector correctly isolated the object of the desired type, built a model, and proceeded to track it.

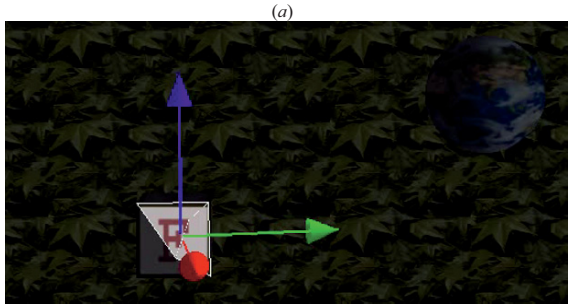
Tracking in these examples proceed despite incomplete OI models. This is particularly evident in Figure 6(b), where incomplete motion segmentation highlights only part of the object. The current motion segmentor has difficulty with image regions containing mostly image gradients perpendicular to the motion. This is because it is segmenting based purely on motion and motion can only be determined along the direction of image gradients. It also segments based on pure 2D translation and, therefore, has difficulty with rotating objects. Despite this issue, an appropriate model is still selected and a partial model built (Figure 6(b)). The object tracker was able to successfully track this OI based on the partial model.

A *real-world* test is shown in Figure 7. In this test, a box was placed on an x/y-table and images were captured using a Sony SNC-RZ30N camera. Once again, the colour IF was tuned to the box and the OI selector correctly separated the box from the background, facilitating model building and subsequent tracking.

Figure 8 shows frames output by the object tracker during tracking, confirming that the model is successfully rebuilt every 31 frames. It tracked the box to within 20 mm positionally and 16° orientationally. Improving the OI model would improve the tracking accuracy.

Figure 4: The motion segmentation map for a sequence containing two objects.

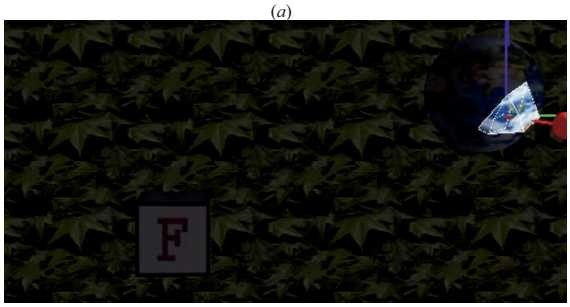
(*a*)



(*b*)

Figure 5: Using the motion segmentation map in Figure 4, the IF framework has been tuned to the cube, resulting in the (*a*) interest map and (*b*) model.

(*a*)



(*b*)

Figure 6: Using the motion segmentation map in Figure 4, the IF framework has been tuned to the globe, resulting in the (*a*) interest map and (*b*) model.

VI. Discussion

There are various possible improvements that could be made to enhance its usability of the novel automated OI selection system presented herein. For example, currently only two IF types have been implemented; additional IF types (such as positional and motion-based IFs) would extend the IF framework's usability. An IF using colour co-occurrence [17] could provide more accurate OI identification than the

(*a*) (*b*) (*c*)

Figure 7: (a) The motion-segmentation map, (*b*) interest map, and (*c*) resulting model for a real-world motion sequence.

(*a*) (*b*) (*c*)

Figure 8: (a) Frame #68 (*a*), (b) #153 (*b*), and (*c*) #237 of the tracking response continuing from Figure 7.

current colour IF. Another possible IF would filter based on similarity to the motion estimated by the object tracker. This would not help OI selection, but would help object separation during subsequent tracking (the model is periodically updated).

The OI selector could also benefit from an improved motion segmentor. The current motion segmentor segments based purely on optical-flow and is 2D only. Using additional information such as colour (e.g., Cucchiara et al. [18]) could improve the object separation. Using higher order flow models, such as affine flow, could yield better segmentation when objects rotate. However, achieving real-time operation with higher-order models could be challenging.

VII. Conclusions

This paper presents a novel and flexible OI selection framework that enables an object tracking system to automatically select which object to track and, subsequently, initiate tracking. A modular and flexible framework of the IFs allows the OI selector to be tailored to identify the objects that the target robotic or surveillance system is interested in.

The selected OI region is passed on to a real-time, 6-dof, pose tracker. Our pre-existing object tracker was extended to enable modelling and tracking of unknown objects within background clutter. The same IF framework used by the OI selector enables real-time separation of the OI from its surroundings, without significant additional overhead.

References

[1] H. de Ruiter and B. Benhabib, "Tracking of rigid bodies for autonomous surveillance," *IEEE Int. Conf. on Mechatronics and Automation*, vol. 2, Niagara Falls, Canada, July 2005, pp. 928–933.

[2] H. de Ruiter and B. Benhabib, "Colour-gradient redundancy for real-time spatial pose tracking in autonomous robot navigation," *Canadian Conf. on Computer and Robotic Vision*, Québec City, Canada, June 2006, pp. 20–28.

[3] H. de Ruiter and B. Benhabib, "On-line modeling for real-time, model-based, 3d pose tracking," in Advances and Innovations in Systems, Computing Sciences and Software Engineering, K. Elleithy, Ed. Springer Netherlands, August 2007, pp. 555–560.

[4] F. Jurie and M. Dhome, "Real time robust template matching," *13th British Machine Vision Conf.*, Cardiff, Wales, 2002, pp. 123–132.

[5] E. Marchand, P. Bouthemy, and F. Chaumette, "A 2d-3d model-based approach to real-time visual tracking," Image and Vision Computing, vol. 19, no. 7, pp. 941–955, November 2001.

[6] T. Drummond and R. Cipolla, "Real-time visual tracking of complex scenes," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 932–946, July 2002.

[7] A. Comport, E. Marchand, and F. Chaumette, "A real-time tracker for markerless augmented reality," *IEEE and ACM Int. Symposium on Mixed and Augmented Reality*, Tokyo, Japan, October 2003, pp. 36–45.

[8] S. Kim and I. Kweon, "Robust model-based 3d object recognition by combining feature matching with tracking," *Int. Conf. on Robotics and Automation*, vol. 2, Taipei, Taiwan, September 2003, pp. 2123–2128.

[9] V. Kyrki and D. Kragic, "Integration of model-based and model-free cues for visual object tracking in 3d," *Int. Conf. on Robotics and Automation*, Barcelona, Spain, April 2005, pp. 1554–1560.

[10] M. Vincze, M. Schlemmer, P. Gemeiner, and M. Ayromlou, "Vision for robotics: A tool for model-based object tracking," IEEE Robotics and Automation Magazine, vol. 12, no. 4, pp. 53–64, December 2005.

[11] Y. Sun and R. Fisher, "Object-based visual attention for computer vision," Artificial Intelligence, vol. 146, no. 1, pp. 77–123, May 2003.

[12] I. Ahrns and H. Neumann, "Space-variant dynamic neural fields for visual attention," *IEEE Conf. on Computer Vision and Pattern Recognition*, Ft. Collins, CO, June 1999, pp. 313–318.

[13] D. Shreiner, Ed., *OpenGL Reference Manual*, 4th ed. Boston, MA: Addison-Wesley, 2004.

[14] C. Everitt. (2006, October) Projective texture mapping. NVidia. [Online]. Available: http://developer.nvidia.com/object/Projective_Texture_Mapping.html.

[15] G. Collins and L. Dennis, "A system for video surveillance and monitoring," *Int. Conf. on Automated Deduction*, Pittsburgh, PA, June 2000, pp. 497–501.

[16] R. Chung, F. Chin, K. Wong, K. Chow, T. Luo, and H. Fung, "Efficient block-based motion segmentation method using motion vector consistency," *IAPR Conference on Machine Vision Applications*, Tsukuba, Japan, 2005, pp. 550–553.

[17] P. Chang and J. Krumm, "Object recognition with color co-occurrence histograms," *IEEE Conference on Computer Vision and Pattern Recognition*, Fort Collins, CO, June 1999, pp. 498–504.

[18] R. Cucchiara, A. Prati, and R. Vezzani, "Real-time motion segmentation from moving cameras," Real-Time Imaging, vol. 10, no. 3, pp. 127–143, June 2004.

The Principles and Planning Process of an Electronic Kanban System

I.A. Kouri, T.J. Salmimaa, I.H. Vilpola
Institute of Industrial Management, Tampere University on Technology
Institute of Human-Centered Technology, Tampere University on Technology
P.O. Box 541 33101 Tampere, Finland

Abstract

The ideas of pull production managed by kanban-cards are well documented in the literature and are in practice worldwide. The advance of ICT technologies like RFID technology presents us with the opportunity to develop more advanced ICT supported electronic kanban solutions. This study defines the factors influencing the success of electronic kanban systems and describes how they should be taken into account in the planning process.

The study suggests that the understanding of original pull planning philosophy is important. The shop floor practices refined in leading companies utilizing the pull production system have to taken into account while planning the practical ICT applications. The UCD process is important, since the original ideas have to be infused in shop floor information and material flow control systems where the human-computer system interaction takes place.

I. INTRODUCTION

The original kanban system is an essential part of Just In Time (JIT) manufacturing or a Lean Production system. It originated in Japanese manufacturing as a simple solution developed to manage repetitive small batch manufacturing. The most recognized and common kanban system is the one that Toyota Motor Company pioneered in the 1950´s. A large amount of literature and reports describe the Toyota kanban system.[1][2] Many companies have developed their own versions of the kanban system to support their specific operations and management principles. The ideas and practices vary from company to company and therefore an exact definition of kanban system practices is difficult.[3] For example, the latest developments like the conwip system are clearly based on the pull production and kanban type control principle, but function differently.[4]

In this study the kanban system theory is based on classic examples of the Toyota Motor Company. The Toyota kanban system and its ideas and background are best documented in the literature and research. Different writers emphasize that the Toyota production philosophy should be applied as a whole system of values and principles.[5]

Development of ICT-technologies, especially the RFID (Radio Frequency Identification Technology) and different factory environment wireless communication systems give new opportunities to develop more automated and sophisticated electronic kanban systems. The basic functions of a kanban system can be realized by utilizing RFID tags to report the material batch status, number of batches and location of each batch. Wireless ICT -technology can be used to report different actions or changes in production batch status. The computer system can calculate system schedules and update kanban parameters based on the real time system status.

Despite the fact that an electronic kanban has many advantages over the traditional card based kanban and is used in the motor vehicle industry, there are very few scientific reports or articles that study or describe the real life operating system and principles of electronic kanban system development and implementation.

This study explains how the original and successful kanban and pull production ideas are best applied in an ICT based electronic kanban system. The study includes a case study that helps to clarify the factors influencing the success of electronic kanban systems and describes how they should be taken into account in planning process of an electronic kanban system.

First the differences between the traditional kanban and electronic kanban are analyzed from a theoretical point of view. Second the factors influencing the success of electronic Kanban design are defined from the point of view of pull production philosophy and user centered design (UCD) theory. The UCD planning process is used in kanban system planning and it's influence into the planning process and design results are reported.

II. THE BASIC ELEMENTS OF THE KANBAN SYSTEM

The kanban system is a pull planning system that authorizes manufacturing based on the state of the production system.[4] The basic idea of the kanban system is quite simple: kanban cards are used as signals to the preceding production phases that authorize a specified component batch delivery or production. The Japanese word kanban can be translated as a visual signboard or card.[6] [4] In practice it is a simple card that specifies a batch to be delivered or manufactured.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 99–104.

The kanban logic

The traditional kanban system described in many sources includes two types of kanban cards. The withdrawal (or transfer or move) kanban authorizes the movement of production batch. This card is sent to the supplier when the material batch consumed from the cycle inventory. It controls the material flow from the supplier to the in downstream manufacturing phase. The production kanban authorizes the production of parts based on withdrawal of materials from the supplier buffer inventory. As the materials are transferred the production kanban card is released and sent in upstream manufacturing phases.

The number of kanban cards limits the number of production batches in operations at the same time. The batch size is calculated carefully based on exact production needs (takt-time), manufacturing setup and cycle time (throughput time).

This two-card system is used when the manufacturing department or supplier are spatially distributed. If the manufacturing operations are located closely in the same area, a production kanban can be used solely to control manufacturing. Many companies have developed their own versions of the kanban system to support their specific operations management needs. [3]

Kanban as a part of JIT production

The prerequisite of the kanban system to operate effectively is the implementation of Just in Time manufacturing principles.

According to Monden [3] the kanban system in the Toyota production system is supported by the following systems:

- Production smoothing
- Standardization of jobs
- Reduction of setup time
- Improvement activities
- Design of Machine layout
- Autonomation

The kanban system is one of the most visible parts of JIT manufacturing, but its applicability is directly interlinked with other JIT manufacturing principles.

One of the most influential characteristics of kanban system is its support to operations improvement. The kanban controls system is used to limit or cut down work in progress inventories, which will reveal the hidden problems in operations. When the problems are made visible they are solved to improve the operations further. After the operations improvement the production batch and batch size are cut down to start the improvement cycle again. The promotion of operations improvement is widely considered to be one of the most important features of a kanban system. [1] [2] [3]

The advantages of the kanban system

According to different authors the pull production system controlled by kanbans has many advantages over push production. The system is easy to understand, highly visible, reliable, interlinks the operations tightly, emphasizes operations and quality improvement and is cheap to implement.[4]

The classic kanban system has however some negative characteristics. It can be applied only to the repetitive production of standard items that can be buffered between manufacturing operations; the material flow, production mix and capacity needed should be stable without major fluctuations, and the quality and reliability of manufacturing operations as well as supply must be high.[4] Many writers emphasize that the company culture and the workers´ attitude and interest in operations improvement are essential for JIT and kanban system success.[4] [7]

III. The electronic kanban applications

Several articles describe the advantages of electronic signals over the card based kanban systems[8]. Some articles notice that electronic kanbans are in use to some extent, but they do not include more detailed information of the principles and practices utilized.[9][10] Motor vehicle manufacturers have been the first companies reported to have electronic kanban systems in place. Rover used as early as 1986 an EDI based signaling system used to pull materials from the suppliers. General Motors was reported in 1990 to have bar-code based pull production system.[11] Toyota had a working interned based kanban solution in their factories in France and England in the year 2000 to pull materials from suppliers.[12] The suppliers attach a bar-code to each delivered container, which is used for batch identification. After the material is used the bar-code is removed from the container.[13][14] Ford has implemented a simple bar-code based pull production system called SMART.[15] BMW is reported to have an electronic kanban system in place.[16] ERP vendors such as SAP have included kanban functionality in their products. The SAP solution utilizes RFID tags attached to containers to automate batch identification and location tracking. The manufacturing decisions can be made based on accurate information of production batch status.[17]

Most of the electronic kanban systems described use RFID-codes or bar-codes in production batch identification.[10] [16] Bar-codes are applied more often to manage material flow between companies, because the containers are used for different batches. RFID tags are more applicable in inter-company situations because the same containers are used many times. The earlier mentioned Toyota example is limited only to pull the materials from the suppliers. Toyota still uses card-based kanbans in its own manufacturing operations.[12]

The advantages of an electronic kanban system

A traditional card based kanban system has some limitations that can be improved with an electronic kanban system. The kanban system itself includes some non-productive work caused by card handling. As the pace of manufacturing operations increases and the production batch size is cut down, the number of card movement increases as well. The cards are lost or misplaced sometimes, which nearly always causes immediate problems in tightly interlinked JIT production. The movement of the kanban card always has some irregularities, because the cards are not moved at the exact time the materials are consumed.[9] The kanban system is difficult to adjust to production mix changes because the cards have to be collected and replaced with new

ones.[18] The kanbans controlling the products that are ramped down have to be replaced with products that are ramped up. The amount of work in progress has to be followed manually to prevent the overproduction of components for down-ramped products.

An electronic kanban system has many advantages over the traditional kanban system. It is as simple as the traditional kanban, but signals are transferred faster. Several writers list the advantages of the electronic kanban system: [9][12][17]

TABLE I
THE ADVANTAGES OF AN ELECTRONIC KANBAN SYSTEM

Removes the problem of lost cards
The demand need is delivered right time
Time and effort needed cards handling is minimized
Fast and effective optimization of kanban cards
Minimizes the material shortages
Improves the supply chain transparency
Helps to analyze the supplier efficiency

The electronic kanban system is more reliable and there are fewer mistakes in kanban card management and decision-making. The computer system can be used to give the production system more visibility, which is a major cornerstone of JIT manufacturing. Increasing visibility gives support to continuous improvement, which is one major factor behind the success of JIT manufacturing.

An electronic kanban system helps to solve the problems related to the traditional kanban system and improves further the speed and quality of operations management. The production mix change is easier to manage. An electronic kanban system can manage the problems related to product-mix change more efficiently, because the location and size of every batch is known and the kanban card change happens automatically within the computer system. The information about operations collected by an electronic kanban system can be utilized in production batch size and throughput time definition. It is possible to automate the calculation of batch sizes and load leveling and scheduling (heijunka). The status of the production system and location and status of the production batch can be monitored and decisions can be made according to the latest information. The electronic kanban system can help to implement a pull production system in a manufacturing environment where the traditional kanban system would face difficulties. An electronic kanban system can be used in production where the product mix changes constantly according to customer needs. The quality problems or machine failures can be included in the computer system logic so that the influence of failures or quality problems are minimized and the recovery is done in a controlled way. Material flow is one of the cornerstones of the JIT production system. An electronic kanban system can bring visibility and improve production and materials management in a layout where the operations are spatially scattered. These ideas are controversial to the traditional JIT ideology which emphasizes the importance of manufacturing and process organization and improvement. However, many companies work in an environment where it is difficult to implement all the JIT manufacturing principles and technologies according to the theory. An electronic kanban, implemented carefully, can possibly work in an environment where a card-based kanban would not function properly.

The design principles of an electronic kanban system

Based on the theory and description of the electronic kanban system, some principles or ideas concerning an electronic kanban system design can be presented. First, the electronic kanban should follow the principle of the traditional card based kanban system. These principles have been developed and tested over a long period of time by world-leading manufacturing companies. These principles include, for example, smoothed and leveled production, mixed model sequencing, stable material flow, operations tight synchronization (takt-time calculations), and pull signals generated by the status of inventory or production system.

Second, the electronic kanban system should support the continuing improvement that is considered by many authors as one of the most powerful features of the kanban system. The traditional kanban system is used to lower down inventories and minimize production batches until hidden problems are revealed. After the problems are corrected, the inventories and batch sizes are reduced to reveal new problems. This improvement approach should be included in the kanban system to obtain most of the advantages of the pull production system. An electronic kanban system should also support the operations improvement by collecting and reporting data about manufacturing operations and material movement and storage.

Third, the system should be user-friendly and the user-system interface should be well designed. The kanban system function is dependent on the users following the principles of the system. The kanban system reliability and ability to manage the production are directly related to the operator reporting the transactions and possible problems in a standardized way. Therefore user involvement in the planning process is important in developing ICT-systems that are applicable in the manufacturing environment. Despite all the possibilities that information technology gives, the system should be as simple as possible from the operator point of view. Production batch status or transactions reporting should be automated whenever possible.

Fourth, the system can be used to solve problems related to the card based kanban system. The production mix change management, process visibility, system speed and reliability improvement are major issues in justifying investment in the electronic kanban system. These functions or needs should be taken into account while planning the software controlling an electronic kanban system.

Fifth, the electronic kanban system will help to overcome manufacturing process shortcomings like machine breakdowns, quality problems or material flow problems. However, the system should in the first place support the operations improvement instead of making it easier to live with the problems.

IV. User-centered design

The principles of User-Centered Design include (1) the active involvement of the users in the design process and a clear understanding of user and task requirements, (2) a multi-disciplinary design, (3) the iteration of the design solutions (4) that are defined by the standard of ISO 13407:1999. The first stage of UCD process is a specification of the context of use that consists of users, users' tasks, equipment, and both the physical and social environment. After analysis the context of use and the user and organizational requirements are specified. As a result of firm understanding of the users and context it is easier to produce the design solutions that can be evaluated against the requirements. The design solution is iterated as long as the solutions fulfill the collected requirements.

It can be proved that the implementation project of any information technology is more effective if the users are satisfied with the system use. In this paper we suggest that User-Centered Design (UCD) process improves the acceptance of the electronic Kanban system on the shop floor. When primary users of a new system (primary users have to be specified) are a part of the design process, both social transparency and the awareness of the electronic kanban advantages are increased among the workers in the production process. By observing the users' tasks, evaluating the use cases of the tasks and reviewing the paper prototypes of electronic kanban interface, the final design solution can be specified on the basis of the requirements of the use context. The major benefit is that the designed solution is evaluated and improved before the final system is specified for a system vendor.

V. The case study

The manufacturing process of the target company includes assembly, pre-assembly, part manufacturing and the raw material production. (Fig. 1) All the materials are assembled together in the pre-assembly phase. The raw material production department produces materials for all the part manufacturing departments. The manufacturing company currently uses a single kanban card system to manage the material flow between the pre-assembly and component manufacturing departments. The component manufacturing schedules and batch sizes are defined based on kanban cards.

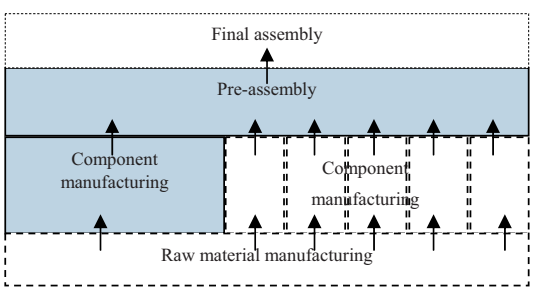


Figure 1. The structure of the manufacturing operations of the case company

Currently, the Kanban cards have the following sequence;

1. Kanban cards are moved from the component production board to container carts holding the components while the component is produced.
2. The truck driver transfers the cart to the pre-assembly station buffer stock
3. The pre-assembly operator takes the card of the cart when the component is taken into production
4. The truck driver takes the card back to board.

Based on daily production mix a relevant set of cards for each type of components is placed in the component production board. The card holds multiple production details. The truck driver receives a weekly paper list of products assigned to each pre-assembly and their schedule. One of the shortcomings of the card-based kanban system is that component scheduling rules are not followed in manufacturing units. The operators and truck drivers schedule their operation in some extend based on their own judgment and experience. The production foremen have guide and control over every operator in some sense. The only way to understand the real time status of operations is to constantly move around the factory and check the situation in different departments.

The card-based kanban system lacks operations transparency, standardized rules of operation and knowledge of real-time production status

The Designing process of an electronic kanban system

The design of an electronic Kanban system has two major challenges: one is the system logic inside the system, and other is the users' logic using the system. In order to have in-depth understanding of the context and production management logic within the current Kanban system, a multidisciplinary team was formed. The aim of the team was to design a system that improves the process, is intuitive to use, is expandable for the other components and processes, and one which users would access into use.

The design process followed the principles and process of user-centered design (UCD) (Fig. 2). The need for UCD was obvious, as there had earlier been a technology pilot incapable of meeting the contextual requirements of the production environment. However, the pilot raised the awareness of the social and functional aspects that the design should cover in addition to the technical details. At first the current card-based Kanban system was carefully studied in order to specify the context of use for the electronic Kanban system. The stakeholders were observed and interviewed, and production data collected. For example, volumes and capacities were checked as well as the worker's tasks and scheduling principles. Multiple points of view were covered. The data was then analysed thoroughly to define the electronic Kanban system specification.

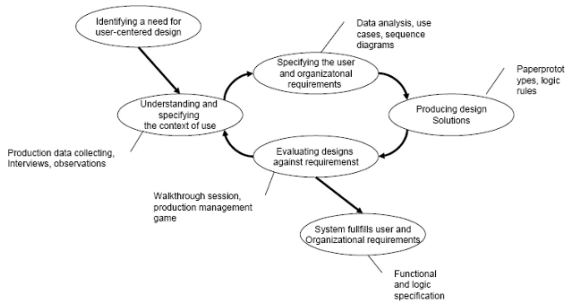


Figure 2. User-centered design process and activities of the case study.

For example, one organizational requirement was to have as little data entering for the system as possible. The data collection should be automated whenever possible. Also the tasks should be scheduled tightly by the system logic, not by the users. Analysis included a value stream map, sequence diagrams and use cases among other descriptive methods, which eased the identification of requirements.

The user interface designs were tested with paper mock-ups and prototypes of system screens. The logic was described with tables of system rules. A key element of UCD is to evaluate the design against the user and organizational requirements. Every user commented on the logic of user interface shown in the paper prototype from the point of view of their role and tasks. The requirement specification established was reviewed and iterated with the design team, which considerably improved the quality and accuracy of the document, as well the logic and functions of the electronic kanban system. The inputs, outputs, and deviations during the task sequence were tested before documentation in the walkthrough session and also in the documentation stage when the final improvements were made.

In this case the logic was evaluated in a production game, which is described in more detail in the next section. The user interfaces and system usage were evaluated in a walk-through session. In the session, all the primary users, i.e. those who directly enter or use the system data, were represented. Use cases were used as a basis for evaluation and the tasks of each user group were covered. The users commented on the system and task design. A few comments also concerned the logic inside the system. Afterwards, all the comments were considered and corrections made.

The production game

The game was used for simulating real-time production. The aim was to test the designed electronic Kanban system logic and its effects to the manufacturing operations.

Every participant had a role (pre-assembly operator, truck driver, component manufacturer, system logic). A rough map of the plant layout formed the basis of the game and toy blocks were used to represent material batches (Fig 3). First everyone was briefed about the processes, set-up and production times. Then the logic and system control were introduced. The game was paused whenever somebody noticed any problem in the game. The cause and effect were briefly analyzed and discussed, and the game was continued again.

Figure 3. The kanban functionality testing game session.

For example, the simulation clearly illustrated that batch size must follow the principles of the mixed model sequencing. On the other hand, the game revealed that if the schedules and batch sizes suggested by the system were followed, then the load of different departments is more levelled and at the same time the number of materials shortages in the assembly is smaller. The game clarified the system logic so well that company management decided that the production workers should play the game to gain a better understanding of the kanban system logic and its interaction to the manufacturing process.

The change from the card-based kanban to the electronic system shifts the decision making from the worker to the system. In return the workers get more real time information on current production status.

Finalizing the design

After a few iterative cycles of UCD, the system structure, functionality and interfaces were specified and designed. The evaluation of the planned system was done with users-to-be a couple of times. As a result a specification document was produced to describe the system for the software and RFID-system vendors.

The document described:

- the context of use, i.e. process, users, their tasks and devices,
- system logic, i.e. rules for calculations
- system configuration, i.e. the parameters needing to be adjusted and their ranges
- functionality, i.e. purpose, data input, process, output, exceptions and errors
- The definition of the most important reports

The process was the first of its kind for many of the participants. The amount of effort put to into the design before implementation raised discussion. On the positive side, every participant had a shared understanding of the potential, capabilities and restrictions included in the system designed.

The organization's capability to use the system diciplined worried most the planning group. However, the user involvement and production game were already identified as tools for promoting for the systematic usage of the electronic Kanban system.

VI. Discussion

The UCD design process aimed at the quality and accuracy of the electronic kanban in helping to follow the four principles. The optimal design solution documented was a result of the design process. The stakeholders who continue implementing can exploit the requirement specification in negotiation with the vendors. The final solution responds to the requirements of the use of context that increase the acceptance of the system. When the social aspect is also included in the design, some challenges related to the changes of the workings habits can be prepared. The behaviour of the users can be forecasted and considered in the system logic and implementation plan. With respect to the UCD design process implemented the case study has achieved its main objectives. The logic of the system and the users' tasks were evaluated and iterated in the multidisciplinary team.

Increasingly, the JIT prerequisites like smoothed and flawless production have to be taken into account in the design of the electronic kanban system. If the current manufacturing process does not support JIT principles, the new system does not necessarily improve the management of operations. The system necessitates commitment to the system rules and operation development. Most challenges which are dependent on the behaviour of people are overcome only with adequate training, that should emphasized in the implementation plan.

Based on our study the UCD design process employed has several advantages over traditional software system project. First active user involvement in the early stage of design supports the requirement specification of the system, because the users can influence the design solutions before implementation of an electronic kanban system. They can directly influence the system features and functions, that improves the functionality of the software system and makes the system implementation easier. Second the change management and the new system acceptance is easier because end-users knowledge of the system principles is better. The change management is based on identified change areas and potential problems found in analysis. Third the multidisciplinary planning team has a better understanding of the users' tasks and the shop floor environment. Participant observation defined the requirements of the use of context and produced a large amount of detailed data of the manufacturing tasks and the current kanban system. Fourth the UCD process itself improved the shop floor operator participation that will improve the new system acceptance.

UCD process itself needs some background preparation and some expertise of the planning tools used. If the background material, such as pre-designs of the system, paper prototypes and production game are prepared carefully the planning process and user involvement is very efficient. In the participant observations only the expert users were observed, which can cause bias in the analysis. Overall, the objectives were well achieved in the UCD design and the company was satisfied with the final result of the design process.

VII. Conclusion

The study suggest that the most of the original kanban ideas should be followed while planning an electronic kanban system. However an electronic kanban system gives possibilities to solve some of the limitations of kanban system, like the model mix change management and failure recovery. The support for continuous improvement should be build into system to achieve the effectiveness of original kanban and JIT-ideas.

The case study suggests that UCD process is effective and efficient tool to be used developing factory floor information systems. The context of use, number of system users and the systems tight integration into factory operations management necessitates careful and multidisciplinary planning process that is difficult to accomplish without user participation to the planning process.

References

[1] Ohno T. 1988 The Toyota Production System, Productivity Press.

[2] Shingo S. 1982 Study of the Toyota Production System, Japan Management Association.

[3] Monden J. 1983, Toyota Production System, Industrial Engineering and Management Press.

[4] Hopp, W., Spearman, M. 2000 Factory physics, Irwin/McGraw-Hill.

[5] Liker J.K. 2004, The Toyota Way, McGraw-Hill,

[6] Schonberger R. 1982, Japanese manufacturing techniques, The Free Press.

[7] Gupta S., Al-Turki Y. Perry R. 1999. Flexible kanban system, International Journal of Operations & Production Management. Vol. 19(10), 1065-1093.

[8] Ansari A., Modarress B. 1995, Wireless kanban, Production and Inventory Management Journal. Vol. 36(1), 60-64.

[9] Drickhamer D. 2005, The kanban e-volution, Material Handling Management, March 24-26.

[10] Vernyi B., Vinas T. 2005, Easing into e-kanban. Industry Week, December, 32.

[11] Harmon R., Peterson L. 1990. Reinventing the factory, The Free Press.

[12] Cullen T. 2002, Toyota speeds parts delivery with e-kanban, Automotive News Europe, Vol. 7(18), s. 49.

[13] Chappell L. 2005. Automotive News. Vol. 79(6153), s. 28.

[14] Kärkkäinen M., Holmström J. 2002. Wireless product identification: enabler for handling efficiency, customization and information sharing, Supply Chain Management, An International Journal. Vol. 7(4), 242-252.

[15] Bicheno J. 2004, The new lean toolbox: towards fast, flexible flow, PICSIE Books.

[16] Kochan A. 2006. BMW innovates at new Leipzig assembly plant, Assembly Automation, Vol. 26(2), 111-114.

[17] SAP 2006, BWP_SB_RFID Technology for Automotive,[http://www.sap.com/industries/automotive/pdf/BWP_SB_RFID_Technology_for_Automotive.pdf].

[18] Mertins K., Lewandrowski U. 1999, Inventory safety stocks of kanban control systems. Production Planning and Control. Vol. 10(6), 520-529.

Design and Control of an Omni-Directional Mobile Robot

Ioan Doroftei, Victor Grosu, Veaceslav Spinu
"Gh. Asachi" Technical University of Iasi, Romania
E-mail: idorofte@mail.tuiasi.ro

Abstract – **This paper introduces an omni-directional mobile robot for educational purposes. The robot has full omni-directional motion capabilities, thanks to its special Mecanum wheels. The paper provides some information about conventional and special wheels designs, mechanical design aspects of the Mecanum wheel and also electronics and control strategies: remote control, line follow, autonomous strategy.**

I. Introduction

Industrial and technical applications of mobile robots are continuously gaining in importance. They are already widely used for surveillance, inspection and transportation tasks. A further emerging market is that of mobile entertainment robots.

One of the main requirements of an autonomous mobile robot is its ability to move through the operational space, avoiding obstacles and finding its way to the next location, in order to perform its task, capabilities known as localization and navigation. In order to know where to go, the robot must have accurate knowledge of its current location. It means, it should use a great variety of sensors, external references and algorithms.

In order to move in tight areas and to avoid obstacles mobile robots should have good mobility and maneuverability. These capabilities mainly depend on the wheels design. Research is continuously going on in this field, to improve the autonomous navigation capability of mobile robotic systems.

This paper introduces an omni-directional mobile robot for educational purposes. The robot has full omni-directional motion capabilities, thanks to its special Mecanum wheels. The present paper provides some information about conventional and special wheels designs, mechanical design aspects of the Mecanum wheel and also electronics and control strategies: remote control, line follow, autonomous strategy. Thanks to its motion capabilities and to its different control possibilities, the robot discussed in this chapter could be used as an interesting educational platform. This report is the result of a research conducted at the Robotics Laboratory of the Mechanical Engineering Faculty, "Gh. Asachi" Technical University of Iasi.

II. Omnidirectional Mobility

The term of omni-directional is used to describe the ability of a system to move instantaneously in any direction from any configuration. Robotic vehicles are often designed for planar motion; they operate on a warehouse floor, road, lake, table etc. In such a two dimensional space, a body has three degrees of freedom. It is capable of translating in both directions and rotating about its center of gravity. Most conventional vehicles however do not have the capability to control every degree of freedom independently.

Conventional wheels are not capable of moving in a direction parallel to their axis. This so called non-holonomic constraint of the wheel prevents vehicles using skid-steering, like a car, from moving perpendicular to its drive direction. While it can generally reach every location and orientation in a 2D space, it can require complicated maneuvers and complex path planning to do so. This is the case for both human operated and robotic vehicles.

When a vehicle has no holonomic constraints, it can travel in every direction under any orientation. This capability is widely known as omni-directional mobility.

Omni-directional vehicles have great advantages over conventional (non-holonomic) platforms, with car-like Ackerman steering or differential drive system, for moving in tight areas [1]. They can crab sideways, turn on the spot, and follow complex trajectories. These robots are capable of easily performing tasks in environments with static and dynamic obstacles and narrow aisles. Such environments are commonly found in factory workshop offices, warehouses, hospitals, etc. Flexible material handling and movement, with real-time control, has become an integral part of modern manufacturing. Automated Guided Vehicles (AGV's) are used extensively in flexible manufacturing systems to move parts and to orient them as required.

The development of an omni-directional vehicle was pursued to further prove the effectiveness of this type of architecture and to add a ground vehicle platform that is capable of exceptional maneuverability. Omni-directional vehicles are divided into two categories that describe the type of wheel arrangement they use for mobility: conventional wheel designs and special wheel design.

III. Wheel Designs

A. Conventional wheel designs

Conventional wheel designs used for mobile robots with omni-directional capabilities can be broken into two types, caster wheels and steering wheels. They have larger load capacities and a higher tolerance for ground irregularities compared to the special wheel configurations. However, due to their non-holonomic nature, they are not truly omni-directional wheels. These designs are not truly omni-directional because when a move with a non-continuous curve is encountered there

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 105–110.

is a finite amount of time before the steering motors can reorient the wheels to match the projected curve [5]. The time constant of this process is assumed much faster than the gross vehicle dynamics for most applications. Therefore, it is assumed to be capable of zero-radius trajectories and retains the term omni-directional. Most platforms that contain conventional wheels and approximate omni-directional mobility incorporate at least two independently steered and independently driven wheels [1]. Active Castor wheels or conventional steered wheels can be used to achieve this near omni-directional mobility.

B. Special wheel designs

Special wheel designs are based on a concept that actives traction in one direction and allow passive motion in another, thus allowing greater flexibility in congested environments [12]. These designs can include the universal wheel, the Mecanum (Swedish) wheel, and the ball wheel mechanism. The universal wheel provides a combination of constrained and unconstrained motion during turning. The mechanism consists of small rollers located around the outer diameter of a wheel to allow for normal wheel rotation, yet be free to roll in the direction parallel to the wheels axis. The wheel is capable of this action because the rollers are mounted perpendicular to the axis of rotation of the wheel. When two or more of these wheels are mounted on a vehicle platform their combined constrained and unconstrained motion allows for omni-directional mobility.

The Mecanum (Swedish) wheel is similar to the universal wheel in design except that its rollers are mounted on angles as shown in Fig. 1. This configuration transmits a portion of the force in the rotational direction of the wheel to a force normal to the direction of the wheel. The platform configuration consists of four wheels located similarly to that of an automobile. The forces due to the direction and speed of each of the four wheels can be summed into a total force vector, which allows for vehicle translation in any direction.

Another special wheel design is the ball wheel mechanism. It uses an active ring driven by a motor and gearbox to transmit power through rollers and via friction to a ball that is capable of rotation in any direction instantaneously.

IV. Mecanum Wheel Design

One of the more common omni-directional wheel designs is that of the Mechanum wheel, invented in 1973 by Bengt Ilon, an engineer with the Swedish company Mecanum AB [8]. The wheel itself consists of a hub carrying a number of free moving rollers angled at 45° about the hub's circumference (see Fig. 1). The angle between rollers axis and central wheel axis could have any value but in the case of conventional Swedish wheel it is 45°. The angled peripheral rollers translate a portion of the force in the rotational direction of the wheel to a force normal to the wheel direction. Depending on each individual wheel direction and speed, the resulting combination of all these forces produce a total force vector in any desired direction thus allowing the platform to move freely in the direction of the resulting force vector, without changing of the wheels themselves.

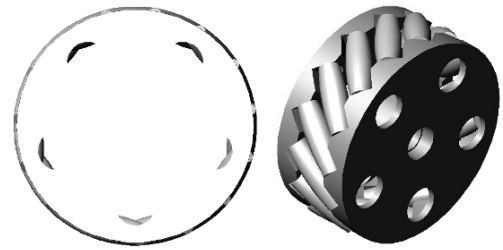


Fig. 1. Mecanum wheel

The rollers are shaped such that the silhouette of the omni-directional wheel is circular (Fig. 1). We can get the shape of a roller if we cat a cylinder, having as diameter the external diameter of the wheel, by a plane angled at γ (the angle between roller and hub axes), in our case $\gamma = 45^\circ$. This shape should respect the equation:

$$\frac{1}{2}x^2 + y^2 - R^2 = 0 . \tag{1}$$

where R is the external radius of the wheel.

If the roller length, L_r, is much smaller than the wheel external radius, R, then the roller shape could be approximated with a circle arc having $2R$ as radius.

In order to get a circular silhouette for the wheel, a minimum number of rollers should be computed (Fig. 2).

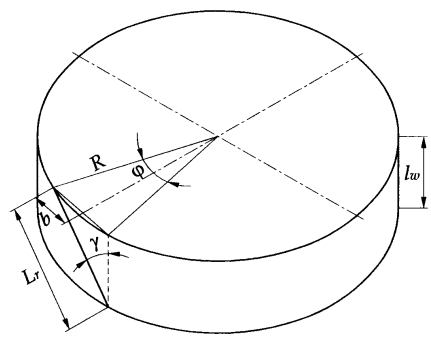


Fig. 2. Wheel parameters

According to Fig. 2, if the roller length is chosen, L_r, we get the number of rollers, n,

$$n = \frac{2\pi}{\varphi}, \tag{2}$$

where:

$$\varphi = 2\arcsin\left(\frac{L_r}{2R\sin\gamma}\right). \tag{3}$$

If we assume that the number of rollers n is known, we can get the roller length:

$$L_r = 2R\frac{\sin\frac{\varphi}{2}}{\sin\gamma} = 2R\frac{\sin\frac{\pi}{n}}{\sin\gamma}. \quad (4)$$

The wheel width will be:

$$l_w = L_r\cos\gamma = 2R\frac{\sin\frac{\pi}{n}}{\tan\gamma}. \quad (5)$$

In our case, $\gamma = 45^\circ$, it means:

$$L_r = 2\sqrt{2}R\sin\frac{\pi}{n}, \quad (6)$$

$$l_w = 2R\sin\frac{\pi}{n}. \quad (7)$$

The rollers are neither actuated nor sensed. The key advantage of this design is that, although the only wheel rotation is powered along the main axis, the wheel can kinematically move with very little friction along many possible trajectories, not just forward and backward [12].

A Swedish omni-directional wheel has 3 DOFs composed of wheel rotation, roller rotation, and rotational slip about the vertical axis passing through the point of contact (see Fig. 3). In the omni-directional wheel, the wheel velocity can be divided into the components in the active direction and in the passive direction. The active component is directed along the axis of the roller in contact with the ground, while the passive one is perpendicular to the roller axis.

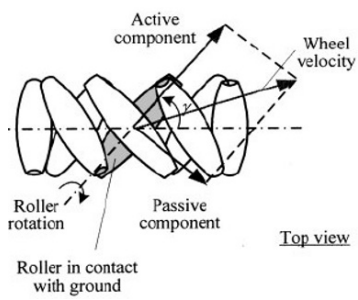


Fig. 3. DOFs in a Mecanum wheel [11]

V. Robot Design

A. Mechanical design

Typical Mecanum-wheel based vehicles have a square or rectangular configuration, with two wheels on each side of the chassis. Using four of these wheels provides omni-directional movement for a vehicle without needing a conventional steering system. In our case, we have chosen a square configuration, in order to simplify the mathematical model and, obviously, the motion control of it (Fig. 4). Our robot is a 450 [mm] long, 382 [mm] wide and 220 [mm] high platform.

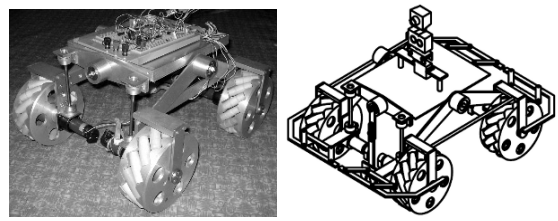


Fig. 4. Omnidirectional robot: (a) photo of the first prototype; (b) CAD design of the second prototype

Each wheel is actuated by its own DC geared MAXON motor. Because the omni-directional capability of the robot depends on each wheel resting firmly on the ground, some are equipped with suspension systems. Even if these designs are for indoor applications (this means they are moving on flat surfaces), having four wheels, they need a suspension system just in case of small waves that could exist on the ground. In our case, a passive suspension system with two spatial four-bar mechanisms ($ABCD$ and $A'B'C'D'$, serial connected) is used, in order to easy adapt the system to the ground (Fig. 5).

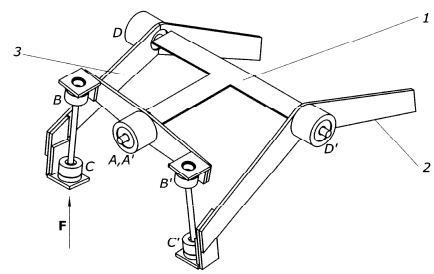


Fig. 5. Suspension mechanism

The robot is able to translate on any direction, forward/backward but also sideways left/right, and turning on the spot, thanks to its special wheels. This is especially helpful when having to maneuver in a tight environment such as a factory floor.

At this time, three solutions for motion control are implemented in the microcontroller:

- Remote control mode;
- Line-follower mode;
- Autonomous mode, thanks to an ultrasonic pair sensors and bumper bars.

Because a single pair of ultrasonic sensors is used, the module is always swinging using a servo and a supplementary gear transmission, mounted on the top of the robot. Even if the servo has an angular stroke of 180°, thanks to the gear

transmission, the sensor is able to scan an angle of 360˚. To avoid obstacle collisions when the robot is going backward for an obstacle avoiding, and also when it run into an obstacle under a small angle, two supplementary pairs of switches have been added (one pair in the front and one in the back).

B. Kinematics

Based on Fig. 6, we can write the robot inverse kinematics as follow:

$$\begin{bmatrix}\omega_1\\ \omega_2\\ \omega_3\\ \omega_4\end{bmatrix}=\frac{1}{R}\begin{bmatrix}1 & 1 & -(l_1+l_2)\\ 1 & -1 & l_1+l_2\\ 1 & -1 & -(l_1+l_2)\\ 1 & 1 & l_1+l_2\end{bmatrix}\cdot\begin{bmatrix}v_x\\ v_y\\ \Omega_z\end{bmatrix} \tag{8}$$

where: ω_i, for $i=1\ldots4$, is the angular velocity of the wheel *i*; R, l_1 and l_2 are some parameters that depend on the robot size; v_x, v_y, Ω_z are the linear and angular velocities of the vehicle.

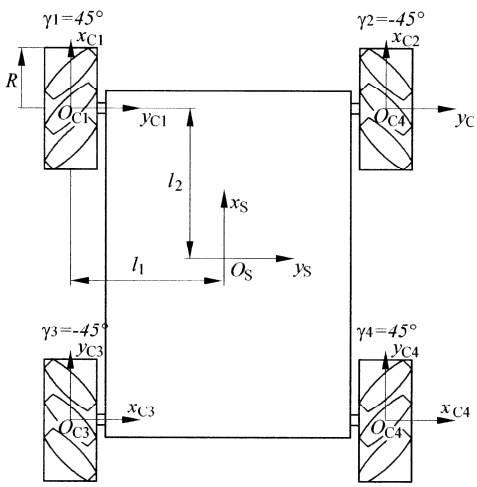


Fig. 6. Planar representation of the vehicle (bottom view)

C. Dynamics

Figure 7 shows the forces that act on the vehicle, where: F_r^j, for $j=1\ldots4$, is the force at roller level, generated by the torque applied to wheel; F_j are the forces applied by the four wheels to the vehicle chassis; F_x and F_y are the resulting forces at the robot center of weight; M_z is a rotationally momentum; F_i^j is the inertial force of each wheel at point of ground contact.

The forces equilibrium at the wheel level can be written as following:

$$\begin{bmatrix}\dot\omega_1\\ \dot\omega_2\\ \dot\omega_3\\ \dot\omega_4\end{bmatrix}=\frac{1}{J_w}\begin{bmatrix}T_1\\ T_2\\ T_3\\ T_4\end{bmatrix}-\frac{\sqrt{2}R}{J_w}\begin{bmatrix}F_1\\ F_2\\ F_3\\ F_4\end{bmatrix} \tag{9}$$

where: T_i is the torque applied to the wheel *i*, $i=1\ldots4$; J_w is the inertia momentum of one wheel.

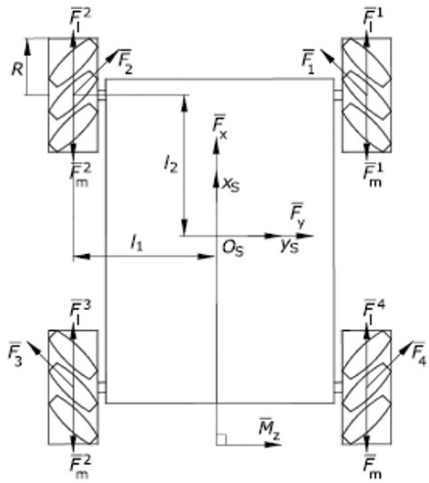


Fig. 7. Force disposal along the vehicle body

(10) describes the forces equilibrium at the chassis level:

$$\begin{bmatrix}F_x\\ F_y\\ M_z\end{bmatrix}=\frac{1}{\sqrt{2}}\begin{bmatrix}-1 & 1 & -1 & 1\\ 1 & 1 & 1 & 1\\ -S & S & S & -S\end{bmatrix}\begin{bmatrix}F_1\\ F_2\\ F_3\\ F_4\end{bmatrix} \tag{10}$$

where $S=\dfrac{l_1+l_2}{\sqrt{l_1^2+l_2^2}}$.

$$\begin{bmatrix}\dot v_x\\ \dot v_y\\ \dot\Omega\end{bmatrix}=\begin{bmatrix}1/m_v & 0 & 0\\ 0 & 1/m_v & 0\\ 0 & 0 & 1/J_v\end{bmatrix}\begin{bmatrix}F_x\\ F_y\\ M_z\end{bmatrix} \tag{11}$$

In (11), m_v is the vehicle mass and J_v is its inertia momentum.

D. Electronics

As we have mentioned before, the robot can be remote controlled, using a command system (Fig. 8), or it can follow a line or to be autonomous, using an ultrasonic sensor for obstacle avoidance.

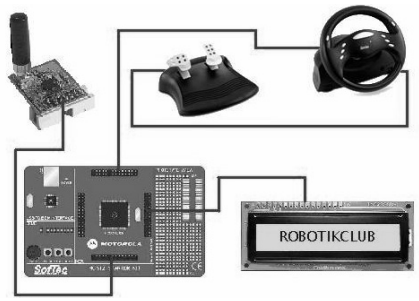


Fig. 8. Command system

In order to receive commands from the system presented in Fig. 8, or to detect obstacles and to drive the motors, an electronics board based on a PIC16F876 microcontroller, and placed on the robot, is used (Fig. 9).

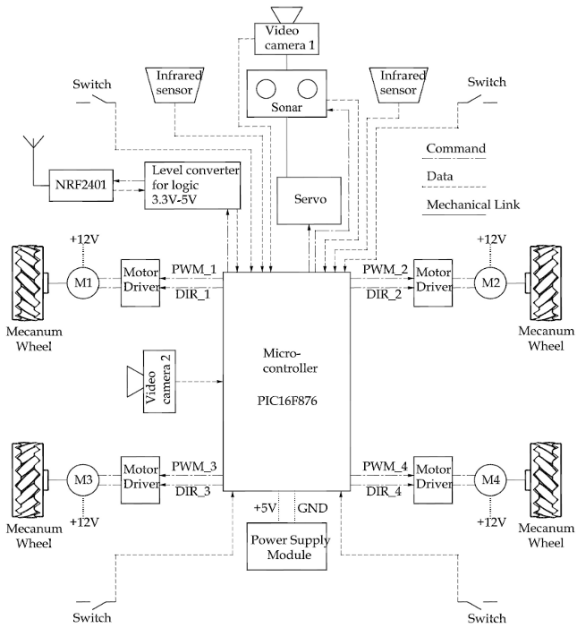


Fig. 9. Electronic board

VI. Robot Control

By "vehicle control" we understand bringing the linear and angular velocities to desired values. In this way, imposing the vehicle velocity we are actually imposing the velocity of each wheel. The above state is valid only when there is no wheel slip present, but for a closed loop control of the platform this strategy is acceptable. To achieve the necessary velocity for each wheel, a PID controller is used. For this purpose we assume that the vehicle is equipped with wheels encoders and accelerometers, used to determine the vehicle speed at each moment of time. For a slip control we have used an Anti-windup like technique, with the difference that a limitation is done to the slip value and not to the command. Hence, the slip is calculated using sensors' data by means of inverse kinematic model.

The aim of this project was to provide the platform with motion control that could be programmed to accommodate various robotic behaviors specified.

A. Line follower mode

Fixed line following is the simplest and most reliable solution, yet is also the most limiting. A physical line is marked on the ground along the path which the robot is to follow [1], [6]. For a robot that is set up in a fixed location for a set task this system is effective but for a research robot with omni-directional capability this approach is seen to be a primitive, though still viable, option.

The main application of this platform being education, line following remain interesting for the students. We have used two OPB704 infrared sensors.

B. Remote control mode

The robot can be remote controlled, using a command system (Fig. 8), and the electronic board (Fig. 9). A steering wheel set is used as a main driving element and two transmitters/receivers based on the NRF2401 circuit, in order to transmit commands to the robot and to receive data from it.

Because the robot can be controlled outside the room where we have the fixed command system, a video camera (video camera 1 in Fig. 9) will transmit all the time images from the robot working space. In such way, we are able to see everything around the vehicle and we can act according to the data received from the video camera.

C. Autonomous mode

To avoid obstacles in autonomous mode, a pair of ultrasonic sensor and two contact switches (two in the front and two in the back) have been used. Because a single pair of ultrasonic sensors is used, the module sonar-camera is always swinging using a servo and a supplementary gear transmission, mounted on the top of the robot. Even if the servo has an angular stroke of 180˚, thanks to the gear transmission, the sensor is able to scan an angle of 360˚. To avoid obstacle collisions when the robot is going backward for an obstacle avoiding, and also when it run into an obstacle under a small angle, two supplementary pairs of switches have been added (one pair in the front and one in the back).

To simplify the control and to reduce the time for the distance measuring procedure, the servo (and ultrasonic sensor) is swinging between two limits ($-\alpha_{\max}$, $-\alpha_{\min}$). The robot velocity along x_S axis depends on $\sum_i D_i p_v$ and its angular velocity depends on $\sum_i D_i p_\omega$, where: D_i is the distance measured to the obstacle, for α_i angular position of the servo

(Fig. 10); p_v and p_w are the computing coefficients for linear and angular speed, respectively.

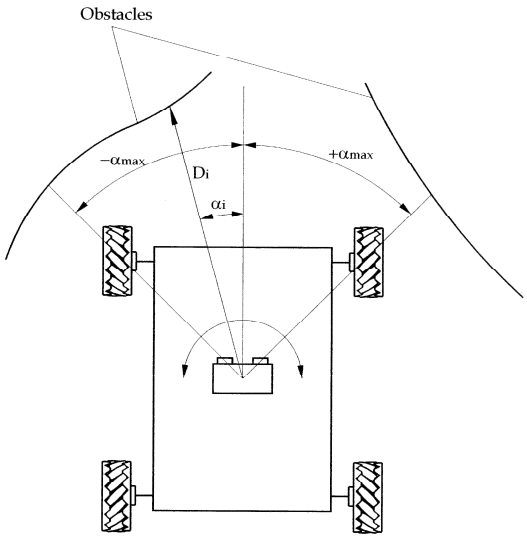


Fig. 10. Obstacle avoidance

If the distance to the obstacle becomes smaller than a minimum imposed value, the linear speed will be zero (the robot stop) and the vehicle will turn, finding an "exit". If this turning motion takes a long time (this time becomes bigger than a maximum imposed value), the servo will swing for the maximum limits (from -180˚ to +180˚) finding a faster solution. This could happen when the robot is entering in a closed space.

If during the movement of the robot it will get an impact with an obstacle (one of the front switches is ON), the vehicle will go backward and will continue turning, and then it will try again to go forward.

The wheel slip is a common problem with the Mecanum wheel, particularly when the robot moves sidewise, as it has only one roller with a single point of ground contact at any one time. To cope with the problem, visual dead-reckoning was used as a slip-resilient sensor [7], [9]. This technique, also used in optical mice, makes use of an on-board video-camera continuously capturing frames of the ground beneath and image processing hardware on the robot determining the speed and direction in which the current frame has moved relative to the previous frame thus allowing the speed and direction of that point of reference to be calculated. For ground image capture, we are using a camera fixed under the chassis of the robot.

VII. Conclusion

Omni-directional vehicles have great advantages over conventional (non-holonomic) platforms, with car-like Ackerman steering or differential drive system, for moving in tight areas. They can crab sideways, turn on the spot, and follow complex trajectories. These robots are capable of easily performing tasks in environments with static and dynamic obstacles and narrow aisles. Such environments are commonly found in factory workshop offices, warehouses, hospitals, etc. Flexible material handling and movement, with real-time control, has become an integral part of modern manufacturing.

This paper introduced an omni-directional mobile robot with Mecanum wheels for educational purposes. The robot has full omni-directional motion capabilities, thanks to its special Mecanum wheels. Some information about conventional and special wheels designs, mechanical design aspects of the Mecanum wheel and also of the robot, as well as electronics and control strategies have been presented.

At this time, three solutions for motion control have been implemented. In order to know the position of the robot according to a reference point, a video camera was used for ground image capture.

Aknowledgment

The authors want to tank to The Romanian National University Research Council for their financial support.

References

[1]J. Borenstein, H.R. Everett and L. Feng, *Navigating Mobile Robots: Sensors and Techniques,* A K Peters, Ltd, MA, USA, 1996.

[2]T. Bräunl, „Eye-Bot: a family of autonomous mobile robots", *Proceedings of 6thInternational Conference on Neural Information Processing*, pp. 645-649, November, 1999, Perth, Australia.

[3]S.L. Dickerson and B.D. Lapin, „Control of an omni-directional robotic vehicle with Mecanum wheels", *National Telesystems Conference Proceedings*, pp.323-328, March 26-27, 1991, Atlanta, USA.

[4]O. Diegel, A. Badve, G. Bright, J. Potgieter and S. Tlatle, „Improved Mecanum Wheel Design for Omni-directional Robots", *Proc. 2002 Australian Conference on Robotics and Automation, Auckland*, 27-29 Nov. 2002, pp. 117-121.

[5]S. Dubowsky, F. Genot, S. Godding, H. Kozono, A. Skwersky, H. Yu, and L. Yu, „PAMM – A Robotic Aid to the Eldery for Mobility Assistance and Monitoring: A Helping-Hand for the Elderly", *IEEE International Conference on Robotics and Automation*, 2000.

[6]H.R. Everett, *Sensors for Mobile Robots: Theory and Application*, A K Peters, Ltd, MA, USA, 1995.

[7]A. Giachetti, M. Campani and V. Torre, "The use of optical flow for road navigation", *IEEE Transactions on Robotics and Automation*, Vol.,14, No.,1, pp.34-48, 1998.

[8]B.E. Ilon, *Wheels for a Course Stable Selfpropelling Vehicle Movable in any Desired Direction on the Ground or Some Other Base*, U.S. Patent, 1975.

[9]K. Nagatani, S. Tachibana, M Sofne and Y. Tanaka, "Improvement of odometry for omnidirectional vehicle using optical flow information", *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2000.

[10] R. Siegwart and I.R. Nourbakash, *Intro to Autonomous Mobile Systems*, MIT Press 2004.

[11] J.B. Song and K.S. Byun, K.S., "Design and Control of a Four-Wheeled Omnidirectional Mobile Robot with Steerable Omnidirectional Wheels", *Journal of Robotic Systems*, 21(4), 2004, pp. 193-208.

[12] H. Yu, S. Dubowsky and A. Skwersky, „Omni-directional Mobility Using Active Split Offset Castors", *Proceedings ASME Design Engineering Technical Conferences*, Baltimore, September 2000.

Preventing Pole-Zero Cancellation for Improved Input Disturbance Rejection in Iterative Feedback Tuning Systems

J.Sikaundi and M.Braae
Department of Electrical Engineering
University of Cape Town
Private Bag, Rondebosch, 7701 South Africa

Abstract **– Modifications to the common IFT algorithm are proposed in order to address the problem of poor internal performance that was noted in instances that are commonly encountered in industrial systems. This is caused by pole-zero cancellation as demonstrated and four proposals that achieve pole-shifting for IFT in cases where it would otherwise produce pole-zero cancellation are investigated.**

I. INTRODUCTION

Iterative feedback tuning (IFT) [1] is a well-studied adaptive control method for obtaining the parameters of a control law without requiring a dynamic model of the process [2].

Its use on industrial problems (with process dynamics that are often dominant first order and controllers that are standard PI) can lead to pole-zero cancellation [3]. This may not always be desirable [4] since, amongst other things, such systems also exhibit poor internal performance [5]. Even though this is particularly important for high-performance controlled systems, methods to avoid pole-zero cancellation in IFT were not found in the literature.

The dynamics of input disturbance rejection (as modelled by a set of transfer functions that incidentally is often overlooked in control engineering texts) provides a practical means of detecting poor internal performance [3]. This observation led to the present investigation in which a number of options that exploit input disturbance rejection are proposed as a means of ensuring that the modified IFT avoids pole-zero cancellation in situations where it might otherwise occur.

Fig. 1 shows the control system configuration that is studied in this paper.

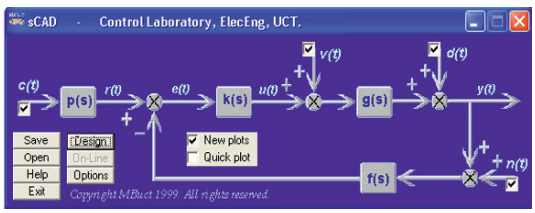


Fig.1. Control Loop under Investigation

The process model is the transfer function g(s) while the controller, feedback element and prefilter are modelled by the transfer functions k(s), f(s) and p(s) respectively. The command signal to the loop is c(t), the input and output disturbances are v(t) and d(t) respectively and signal n(t) approximates the sensor noise. The error signal is e(t) and serves as an alternate signal to y(t) in judging loop performance. In this paper the feedback element f(s) and the prefilter p(s) are unity since the primary focus is on the use of IFT in an industrial setting where k(s) is more commonly perceived as the only controller in the loop.

The twelve transfer functions that map external signals (c, d, v and n) that enter the closed loop to the responses (y, e and u) of the closed loop system are:

$$\begin{bmatrix} y \\ e \\ u \end{bmatrix} = \begin{bmatrix} \frac{n_p n_g n_k d_f}{d_p(d_g d_k d_f + n_g n_k n_f)} & \frac{d_g d_k d_f}{d_g d_k d_f + n_g n_k n_f} & \frac{d_k d_f n_g}{d_g d_k d_f + n_g n_k n_f} & \frac{-n_g n_k n_f}{d_g d_k d_f + n_g n_k n_f} \\ \frac{n_p d_f d_g d_k}{d_p(d_g d_k d_f + n_g n_k n_f)} & \frac{-n_f d_g d_k}{d_g d_k d_f + n_g n_k n_f} & \frac{-n_f d_k n_g}{d_g d_k d_f + n_g n_k n_f} & \frac{-n_f d_g d_k}{d_g d_k d_f + n_g n_k n_f} \\ \frac{n_p n_k d_f d_g}{d_p(d_g d_k d_f + n_g n_k n_f)} & \frac{-n_k n_f d_g}{d_g d_k d_f + n_g n_k n_f} & \frac{-n_k n_f n_g}{d_g d_k d_f + n_g n_k n_f} & \frac{-n_k d_g n_f}{d_g d_k d_f + n_g n_k n_f} \end{bmatrix} \begin{bmatrix} c \\ d \\ v \\ n \end{bmatrix}$$

where each transfer function has been expressed as a ratio of polynomials (e.g. $g(s) = \frac{n_g}{d_g}$). For ease of reference these transfer functions can be represented as the matrix equation:

$$\begin{bmatrix} y \\ e \\ u \end{bmatrix} = \begin{bmatrix} h_{yc} & h_{yd} & h_{yv} & h_{yn} \\ h_{ec} & h_{ed} & h_{ev} & h_{en} \\ h_{uc} & h_{nd} & h_{uv} & h_{un} \end{bmatrix} \begin{bmatrix} c \\ d \\ v \\ n \end{bmatrix} \quad (1)$$

where symbol "h" indicates transfer functions of the closed loop system.

Thus the first two rows of the transfer function matrix represent the system performance whilst the third row represents the cost of achieving that performance.

Some of these transfer functions are similar so in general the designer would consider nine transfer functions, namely, h_{yc}, h_{yd}, $h_{yv,}$ h_{yn} (or h_{uv}), h_{ec}, h_{ed} (or h_{en}), $h_{ev,}$ $h_{uc,}$ h_{ud} (or h_{un}). For optimal performance with reasonable cost, these transfer functions of the closed loop system must be designed with care and are likely to involve trade-offs specific to the problem under consideration [10].

Inspection of these nine transfer functions show that transfer functions h_{yv} and h_{ev} will retain the poles of the process model g(s) when pole-zero cancellation occurs (i.e. when $n_k = d_g$). If these poles of the plant are slow then there will be a slow input disturbance rejection. The slow poles can be cancelled from transfer function h_{ev} using $n_f = d_g$. Note

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 111–116.

however that the poles of g(s) cannot be cancelled using p(s), f(s) or k(s) in the case of transfer function h_{yv}. Therefore the characteristic equation for the rejection of input disturbances, v(t), can only be improved using pole-shifting. This observation motivates the current research and emphasizes the significance of studying input disturbance rejection.

Section II and III of this paper illustrate the effect that pole-zero cancellation has on the rejection of input disturbance and show how pole-zero cancellation readily arises in industrial situations. Section IV considers four pole-shifting methods that have been researched before conclusions are made on the current status of the work.

II. THE PROBLEM OF INPUT DISTURBANCES

Iterative Feed Tuning optimizes an appropriate cost function for the loop under consideration, such as the well-known error squared function:

$$J = \frac{1}{2N}\sum_{i=1}^{N} e_i^2 + \lambda u_i^2 \qquad (2)$$

where e in this case is the difference between the output of the controlled system and that of a desired model (i.e. e = y – ym) whilst u is the control signal. The IFT algorithm [1], then tunes the parameters of its limited-complexity controller to minimize this cost function.

If the optimization is performed for setpoint tracking with no input disturbances and the cost function is at its minimum when the poles of the plant are cancelled then the poles of the plant will be cancelled by the IFT algorithm. If a non-periodic, unexpected input disturbance occurs under this condition then the plant may never recover (or take an unacceptably long time to recover).

For example consider the process model $g(s) = \frac{A}{s(1+Ts)}$, with a controller of the form $k(s) = \frac{k(1+Ts)}{c_o + c_1 s}$. The resulting closed loop system will not recover from an input disturbance, as shown in Fig. 2 for the plant $g(s) = \frac{1}{s(1+5s)}$ and controller $k(s) = \frac{0.5(1+5s)}{1+0.1s}$. A unit step change was applied to the setpoint at time t=0[s] and to the input disturbance at time t=30[s].

Clearly this is unacceptable. If pole-shifting rather than pole-zero cancellation had occurred, then the characteristic equation of all the transfer functions would have been the same for both setpoint change and input disturbance (although it should be mentioned that transfer functions h_{yc}, h_{ec} and h_{uc} can be manipulated by p(s) when this block is used.)

The decision to cancel slow stable poles must be a design choice [4, 6] since this may be advantageous in other situations. The focus of this research is to propose and investigate pole-shifting versions of IFT that could be used when avoidance of pole-zero cancellation is important.

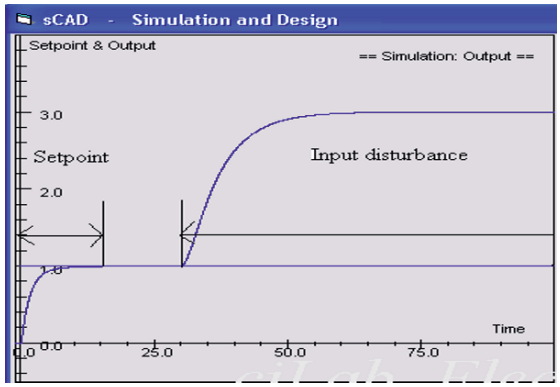


Fig.2. An example of poor input disturbance rejection

III. POLE-ZERO CANCELLATION IN IFT OF A COMMON INDUSTRIAL PROCESS

A common industrial situation is considered in which the process dynamics are first order, $g(s) = \frac{A}{1+sT_p}$, and the controller takes the form of a traditional PI controller, $k(s) = \frac{b_1 s + b_0}{s}$ with two tuneable parameters. The desired model for the closed loop is taken as being dominant first order $m(s) = \frac{1}{1+sT_m}$. It is shown below that the controller that gives a minimum of this cost function (with $\lambda = 0$) is one that has zeros at the same place as the poles of the process. This was confirmed by simulation in [3].

The closed loop transfer function equals that of the model for setpoint tracking when $h_{yc} = \frac{kg}{1+kg} = m$

Rearranging this equation gives:

$$k = \frac{1}{g} \times \frac{m}{1-m}$$
$$= \frac{d_g}{n_g} \times \frac{n_m}{d_m - n_m} \qquad (3)$$

The optimal controller for (2) can only be obtained if it belongs to the class of controllers being tuned. For the above case study (3) gives:

$$k = \frac{1}{AT_m} \times \frac{1+sT_p}{s} \qquad (4)$$

This controller belongs to the class of PI controllers. Therefore IFT performed on this case study will result in pole-zero cancellation.

The reasons for choosing a desired first order closed loop model, m(s), is that it is simple (and thus readily understood by plant personnel) and it reduces or eliminates the need to

specify an explicit limit on the control law because the plant input u is implicitly defined by the model since: $u = \frac{m}{g} r$, [3].

The response of the closed loop system with pole-zero cancellation is shown in the Fig. 3. Here the assumed process model was $g(s) = \frac{1}{1+10s}$ and desired closed loop model was $m(s) = \frac{1}{1+2s}$. The standard IFT algorithm tuned the controller parameters to $k(s) = \frac{0.5(1+10s)}{s}$.

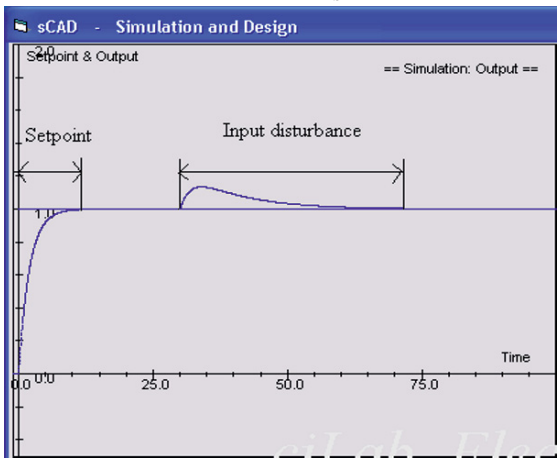


Fig.3. Poor internal performance

Clearly the input disturbance imposed at time, t=30[s], takes a longer time to be rejected than the output takes to reach a setpoint, changed at time t=0[s].

IV. PREVENTING POLE-ZERO CANCELLATION

From the above observations it was proposed [3] that IFT be done through the input disturbances rather than the setpoint. This will be the basis of the procedures presented below. As illustrated above pole-zero cancellation can give a closed loop with unacceptable input disturbance rejection. Therefore if the loop is tuned for rejection of input disturbances rather than setpoint tracking then pole-shifting rather than pole-zero cancellation may be achieved. The equations used in this modification of the IFT algorithm are shown below.

$$y = \frac{g}{1+kg} v = Vv\,,\ e_m = y - y_m \tag{5}$$

Using the cost function in (1) with λ = 0

$$\frac{\partial J}{\partial \rho} = \frac{1}{2N} \sum_{i=1}^{N} 2em\left(\frac{\partial y}{\partial \rho} - \frac{\partial ym}{\partial \rho}\right)$$

$$= \frac{1}{2N} \sum_{i=1}^{N} 2em\left(\frac{\partial y}{\partial \rho}\right)$$

$$= \frac{1}{2N} \sum_{i=1}^{N} -2em \frac{\partial k}{\partial \rho}(V^2 v) \tag{6}$$

Under ideal conditions an input disturbance would not affect the output at all (i.e. $y_m = 0$ when the setpoint is zero). The problem with this requirement is that the gain of the controller grows without bounds as shown in [3].

The same results can be obtained by doing classical IFT [1] through the setpoint with the following conditions: Constant λ in (2) is zero, the setpoint is constant and a unit step input disturbance is imposed during experiment 1. Experiment 2 is left the same. This is because the two methods are mathematically equivalent [7]. The gain of the control law can be limited by using a suitable non-zero value for λ as reported by [8].

Four methods are investigated in this contribution:

1. Having a model for input disturbance rejection,
2. Using time weightings in the cost function,
3. Using a model with overshoot since it is shown in [6] that whenever there is pole-shifting there will be overshoot, and
4. Approximate pole placement IFT.

These proposals will be tested using a first order plant and a PI controller because these are widely encountered in industry and it was shown that pole-zero cancellation will occur with IFT when there are no disturbances present in the system.

A. Using an Input Disturbance Rejection Model

A logical alternative to penalizing the control effort directly in the cost function through non-zero λ values is to use a desired model for input disturbance rejection. This would also constrain the control effort as indicated above and in [3].

The problem moves to the choice of a model for the rejection of input disturbances. If pole-zero cancellation does not occur then the characteristic equations of the input disturbance transfer function and the other loop transfer functions in (1) will be the same. Therefore an equivalent input disturbance rejection model with the same characteristic equation as the desired setpoint tracking transfer function can be used. With reference to previous experience an input disturbance rejection model with a first order characteristic equation of the form $m(s) = \frac{ds}{1+T_m s}$ is attempted. It contains two parameters that the user needs to specify.

The actual model chosen for experimentation is $m(s) = \frac{1.143s}{1+1.143s}$. Its time constant ensures that its transients settle to within 0.001 of the initial value in 8 seconds (i.e. $7T_m$

=8). Table 1 shows how this modified IFT positions closed loop poles for four different processes with λ = 0 in the cost function (2).

TABLE 1
EFFECT OF 1ST ORDER MODEL ON CLOSED LOOP POLE POSITIONS

g(s)	Closed loop pole positions	Damping factor	Settling Time (7T)
$\frac{1}{1+2s}$	-0.534+/-j0.783	0.563	13.1
$\frac{1}{1+3s}$	-0.418+/-j0.726	0.499	16.7
$\frac{1}{1+4s}$	-0.352+/-j 0.685	0.457	19.9
$\frac{1}{1+5s}$	-0.307+/-j0.653	0.425	22.8

A more realistic model for the rejection of input disturbance would allow responses that start at zero and grow to a maximum before returning to zero, as shown in Fig. 3. This implies a characteristic equation that is second order or higher. Therefore the second model chosen is one with the same settling time as the previous model (and a high damping factor): $m(s) = \frac{cs}{s^2 + 1.75s + 1}$ where c is a factor that can be adjusted to limit the maximum value of the response to the disturbance. For example a value of 0.625 gives a peak of 0.250 times the size of the disturbance. The characteristic equation ensures a fast response i.e. Settles to a 0.001 band within 8 seconds. Different values of c gave different closed loop pole positions as shown in Table 2 for the process $g(s) = \frac{1}{1+5s}$.

TABLE 2
EFFECT OF c ON CLOSED LOOP POLE POSITIONS

Maximum height as a fraction of the input disturbance	C	Closed loop poles	Damping factor	Settling time (7T)
0.25	0.625	-0.348+/-j0.608	0.497	20.1
0.20	0.500	-0.407+/-j0.623	0.547	17.2
0.15	0.375	-0.506+/-j0.632	0.625	13.8
0.10	0.250	-0.716+/-j0.584	0.775	9.8
0.05	0.125	-0.962+/-j0.675	0.819	7.3

This method avoids pole-zero cancellation but may lead to lightly damped poles. It also shows that the values of c affect the settling time: As c decreases, settling time decreases whilst the damping factor increases. However it did prevent the gain of the controller from increasing to infinity.

B. Time Weighted IFT

This method is based on the normal time-weighted IFT that is mentioned in [1, 9]. It takes the focus off the transient phase of the response. A problem with time-weightings is that it may lead to an oscillatory system [9] which may be unacceptable for some applications. Nonetheless it is now implemented by perturbation of the input disturbance to avoid pole-zero cancellation, rather than by changes in the setpoint (as used in other work).

According to [9] the simple way of doing time-weighted IFT is by making the weighting on the control law one throughout the experiment and that on the model tracking factor zero during the desired transient time and one afterwards. The cost function that is to be optimized is:

$$J = \frac{1}{2N}\sum_{i=1}^{N}(w_e e_i^2 + w_u \lambda u_i^2) \tag{7}$$

where w_e and w_u are the time weightings on u and e respectively.

$$\frac{\partial J}{\partial \rho} = \frac{1}{2N}\sum_{i=1}^{N} 2em\left(\frac{\partial y}{\partial \rho} - \frac{\partial ym}{\partial \rho}\right) + 2\lambda u \frac{\partial u}{\partial \rho} \tag{8}$$

$$= \frac{1}{2N}\sum_{i=1}^{N} w_e\left(-2em\frac{\partial k}{\partial \rho}(V^2 v)\right) + w_u\left(2\lambda\left(\frac{\partial k}{\partial \rho}(Vv) - k\frac{\partial k}{\partial \rho}(V^2 v)\right)\right)$$

$$= \frac{1}{2N}\sum_{i=1}^{N} w_e\left(-2em\frac{\partial k}{\partial \rho}(y_2)\right) + w_u\left(2\lambda\left(\frac{\partial k}{\partial \rho}(y_1) - k\frac{\partial k}{\partial \rho}(y_2)\right)\right)$$

The IFT algorithm defined by the above equations was applied to the process modelled by $g(s) = \frac{1}{1+5s}$ and a disturbance rejection time of 8 seconds to its 0.001 band that allows the results to be compared with those presented earlier. The resultant closed loop pole positions for different values of λ are shown in the Table 3.

As can be seen the damping factor of the closed loop system varies nonlinearly as λ increases. The settling time of the closed loop system increases as λ increases, as expected intuitively. (The damping factor can also depend on the required settling time [9]).

C. Using a Model with Overshoot

It was suggested in [4] that pole-shifting also results in overshoot when the setpoint is stepped. Therefore it is surmised that a model with overshoot may encourage pole-shifting. Clearly this model must have less than the maximum overshoot allowed and requires a second order model.

TABLE 3
EFFECT OF LAMBDA ON CLOSED LOOP POLE POSITIONS

λ	Closed loop poles	Damping factor	Settling time (7T)
0.000	-0.601+/- j0.360	0.858	11.6
0.001	-0.334+/-j1.188	0.278	21.0
0.010	-0.237+/-j0.310	0.607	29.5
0.050	-0.127+/-0.273	0.422	55.1
0.100	-0.072+/-j0.299	0.234	97.2

The setpoint tracking model that was chosen is $m(s) = \frac{1}{s^2 + 1.75s + 1}$. This has the same characteristic equation as the model in section A. Classical IFT [1], with λ in (2) at zero, was done using this model. The results are summarized in the Table 4.

TABLE 4
EFFECT OF MODEL WITH OVERSHOOT ON DIFFERENT PLANTS

g(s)	Closed loop poles	Damping factor	Settling time
$\frac{1}{1+2s}$	-0.430+/-j0.362	0.765	16.279
$\frac{1}{1+3s}$	-0.375+/-j0.278	0.803	18.667
$\frac{1}{1+4s}$	-0.340+/-j0.212	0.849	20.588
$\frac{1}{1+5s}$	- 0.332+/-j0.155	0.906	21.084

This technique gave pole shifting as opposed to pole-zero cancellation. This meant that the characteristic equations of all transfer functions are the same. However they were all different from that of the model that had poles at 0.875+/-j0.484 and a damping factor of 0.875.

D. Approximate Pole Placement IFT

The aim of this method is to use pole placement to avoid pole-zero cancellation and thus ensure that the characteristic equations of all the transfer functions are the same. The basic idea is that the zero of the model, m(s), is continuously set to the same value as the zero of the controller and that the low-frequency gain of the model must be one.

With this case study, pole placement can be approximated in the following manner. The closed loop transfer function is

$$h_{yc} = \frac{A(b0 + b1s)}{Ts^2 + (Ab1 + 1)s + Ab0}$$

$$= \frac{\frac{Ab1s}{T} + \frac{Ab0}{T}}{s^2 + \frac{(Ab1 + 1)s}{T} + \frac{Ab0}{T}} \quad (9)$$

A pole placement design will place the poles of the closed loop at a location defined by a characteristic equation such as s^2+cs+d. However the process model (i.e. A and T) is not known so it will not be possible to calculate where the zeros of the model for a particular characteristic equation must be. Comparing the characteristic equation of h_{yc} to the desired closed loop response, the constants are c = $\frac{(Ab1 + 1)}{T}$ and d = $\frac{Ab0}{T}$. Therefore $\frac{A}{T} = \frac{d}{b0}$ and the model will be $\frac{d}{b0} \times \frac{b1s + b0}{s^2 + cs + d}$. This specifies the desired characteristic equation through its denominator while the zeros are located by the evolving zeros of the PI controller. This model always ensures that the final gain is one and the model is continuously changing as the IFT algorithm alters the controller. The zero of the controller is always present in the model so the controller zero will not attempt to cancel the plant poles – as deduced from (3). This is because it would also be cancelled, making the resulting closed loop even more different from the model. Pole placement occurs because the only difference between the model and the closed loop is the characteristic equation.

The equivalent input disturbance rejection model will be $h_{yv} = \frac{d}{b0} \times \frac{s}{s^2 + cs + d}$. This also depends on the controller parameter b0. As a result it will have to be included into the differential equation to calculate the minimum of the cost function. i.e.

$$\frac{\partial em^2}{\partial \rho} = 2 \times em \times (\frac{\partial y}{\partial \rho} - \frac{\partial ym}{\partial \rho}) \quad (10)$$

where $\frac{\partial y}{\partial \rho}$ is as defined earlier and

$$\frac{\partial ym}{\partial b0} = -\frac{d}{b0^2} \frac{s}{s^2 + cs + d} \times v \quad (11)$$

$$\frac{\partial ym}{\partial b1} = 0$$

This method was applied to both stable and unstable plants with equal success. Pole placement IFT was done using the characteristic equation s^2+2s+1.25. The results are shown in Table 5. The initial controller was selected to be $\frac{1+1s}{s}$ for stable plants and $\frac{-3-3s}{s}$ for unstable plants because IFT is to be applied on a stable closed loop [1].

TABLE 5
IFT POLE PLACEMENT

g(s)	g(s) poles	Final k(s) zeros	Closed loop h(s) poles
$\frac{1}{1-6s}$	0.167	-0.571	-1.03 +/-j0.46
$\frac{1}{1-5s}$	0.200	-0.563	- 1.03 +/-j0.46
$\frac{1}{1-4s}$	0.250	-0.550	- 1.03 +/-j0.46
$\frac{1}{1-3s}$	0.333	-0.533	- 1.03 +/-j0.46
$\frac{1}{1-2s}$	0.500	-0.497	-1.03 +/-j0.46
$\frac{1}{1+2s}$	-0.500	-0.811	-1.04+/-j0.45
$\frac{1}{1+3s}$	-0.333	-0.736	-1.04+/-j0.45
$\frac{1}{1+4s}$	-0.250	-0.703	- 1.04+/- j0.45
$\frac{1}{1+5s}$	-0.200	-0.682	-1.04+/- j0.45
$\frac{1}{1+6s}$	-0.167	-0.671	-1.04 +/- j0.45

It appears that the closed loop poles remain fixed irrespective of the open loop pole positions, as expected. There is pole placement and the characteristic equation is similar to that of the model.

E. Generalization of Approximate Pole Placement in IFT

For higher order processes the controller has to be a higher order. For example second order plants need controllers with three variables, i.e. $k(s) = \frac{b2s^2 + b1s + b0}{s(0.01s+1)}$ in which non-dominant poles might have been used to ensure causality.

It is found that using this method for a first order plant the differential of the model ($\frac{\partial ym}{\partial b0}$) is not required. On the other hand when doing pole placement for a second order plant $\frac{\partial ym}{\partial b0}$ is required for the pole placement algorithm to work.

The need to insert a non-dominant pole in the controller for pole placement in a second order plant necessitates a faster sampling rate for the system. Also the tuning of the high frequency variable b2 takes a long time. This may be due to step inputs not having enough high frequency excitation for tuning this parameter.

CONCLUSION

The traditional IFT algorithm has been found to produce control laws with poor internal performance for systems that unfortunately are encountered often in industry. Four methods that address this problem have been proposed and evaluated. Three of these methods are based on models of disturbance rejection and provide a more direct approach to specifying system dynamics than those that penalise the control effort. Of these modified IFT methods the last one proposes a novel adaptive IFT that was found to prevent pole-zero cancellation. Thus the approximate pole placement method is the best of the techniques investigated. The reason is that the closed loop speed of response and disturbance rejection can be specified directly by choosing the pole positions of the model.

ACKNOWLEGMENT

The financial assistance of the Department of Labour (DST) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the DST.

REFERENCES

[1] H. Hjalmarsson, M. Gevers, S. Gunnarson and 0. Lequin, “Iterative feedback tuning: theory and applications”, IEEE Control Systems, August , 1998, p26-41
[2] H. Hjalmarsson, M. Gevers, “Special section on algorithms and applications of iterative feedback tuning,” Control Engineering Practice, Volume 11, Issue 9, September 2003, Page 1021
[3] H-Y. Wei, S.I. Ginsberg and M. Braae, “Engineering an iterative feedback controller”. Proc.SACAC Control Conf, p64-69, July 2006.
[4] S.F. Graebe and R.H. Middleton, “Stable open loop poles: to cancel or not to cancel,” Proc 34th Conf. Decision & Control, New Orleans, December 1995, p311-316.
[5] P.A. Laplant, “Comprehensive dictionary of electrical engineering,” CRC IEEE Press, 1999.
[6] R.H. Midleton and S.F. Graebe, “Slow stable open-loop poles: to cancel or not to cancel”. Automatica, 35, p877-886, 1999.
[7] H.M. Prochazka, M. Gevers, B.D.O. Anderson, F. Christel, “Iterative feedback tuning for robust controller design and optimization,” [Online] Available http://www.inma.ucl.ac.be/~gevers/PublisMig/S7.pdf [2007 August 16]
[8] J. Sobota, M. Schlegal, “Iterative feedback tuning of a PID controller,” [Online]. Available http://www.rexcontrols.com/_new/downloads/clankyt/ift_iftarticle.pdf [2007 June 22]
[9] M. Gevers, “A decade of progress in the iterative process control design: from theory to practice,” Journal of Process Control 12 (2002), p519-531.
[10] M. Rapson, “Pareto analysis of controller design methodologies for integrator plus dead time processes,” Eurocon 2007, p2607-2606, 2007.

General Inverse Neural Current Control for Buck Converter

José Guillermo Guarnizo M.[1] Javier Antonio Guacaneme M.[2] Cesar Leonardo Trujillo R.[3]
jgguarnizom@unal.edu.co *jguacaneme@udistrital.edu.co* *cltrujillo@udistrital.edu.co*

LABORATORIO DE INVESTIGACION EN FUENTES ALTERNATIVAS DE ENERGIA LIFAE
UNIVERSIDAD DISTRITAL FRANCISCO JOSE DE CALDAS

Abstract

The neural networks have been used in different applications of automatic control systems, many of them in converters of continuous voltage. Different works have been presented where neuronal controllers are used in Buck converters for voltage control, but works where this technique is used for the current control of these devices to a rated load, remain unknown; It was made a search to know the state-of-the-art of this thematic, there was not found applications regarding to this issue. In this simulation, a generalized inverse neuronal control is used, in order to control de current in a Buck Converter. Afterwards, to improve the performance of the controller, it is used a controller of neuronal modules in parallel, trained and activated by unit load regions, this is proposed in order to evaluate their viability to be used this type of neural control, in a posterior implementation of Hardware.

In the present article a controller of these characteristics in a Buck converter has been simulated, with output current of 8A, and a nominal charge of 1.5 Ω and lower impedances, using inverse neural controller.

1. Introduction

The neural network has been used for applications in control systems, due to multiple advantages nowadays, which create great flexibility for nonlinear systems [1], these systems are adapted to disruptions [2], and a very important issue, it is unnecessary a previously mathematical knowledge of the controller system, only input – output patrons [3].

A kind of neural network control is the inverse neural control [4], it consists on taking input – output dates, and training the neural control where the input of the system is the output of the neural controller, and viceversa, the neural network will give the input needed for an output system when it is required.

There are many publications using neural networks for DC-DC converter control voltage systems [5],[6], [7], but works and publications about current control of Buck converter using Neural Control were not found; to start an investigation about this, its necessary to present simulations of the inverse neural control for current in a Buck converter, using Simulink of Matlab.

The summary of this article is presented as follows. The section 2 presents theory about the Buck converter. Section 3 shows an introduction of neural networks and the inverse neural control. Section 4 presents the simulations and results of the different cases looking for an inverse neural current control. Section 5 discusses conclusions, viability of hardware implementation, and future works.

2. Buck Converter

The commuted regulators are based on how to handle the time in which the initial tension is regulated, this is made through a commuted device, usually a semiconductor (Mosfet, IGBT) and an energy storage device of inductive character depending on the topology of the regulator, the output tension could be higher, lower or inverse [8].

[1] José Guillermo Guarnizo Marín: Electronic Engineer Universidad Distrital Francisco Jose de Caldas, Student of Master in Industrial Automatization Universidad Nacional de Colombia, member of LIFAE, Student Member IEEE.
[2] Javier Antonio Guacaneme Moreno: Electronics Engineer, Professor Universidad Distrital Francisco Jose de Caldas, Master in Electrical Engineer Universidad Nacional de Colombia, member of LIFAE, Member IEEE
[3] Cesar Leonardo Trujillo Rodriguez: Electronics Engineer, Professor Universidad Distrital Francisco Jose de Caldas, Master in Electrical Engineer Universidad Nacional de Colombia, member of LIFAE, Member IEEE

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 117–122.

It is necessary to take into account 2 times in the commutation of the commuted device, it is said that the period of the signal is defined as follows:

$$T = T_1 + T_2 \quad (1)$$

In a time T_1 a commutation device permits a connection through an input tension and the output system. For the time T_2 the device of the commutation will interrupt the connection; that exists between power supply and energy storage device, depending on the topology, in a moment it will transfer energy and in another it will storage energy, and the output tension will be:

$$Vo = \frac{Vi * T_1 + 0 * T_2}{T} = Vi * \frac{T_1}{T} \quad (2)$$

Where:

$$\delta = \frac{T_1}{T} \quad (3)$$

Being δ the duty cycle of the signal.

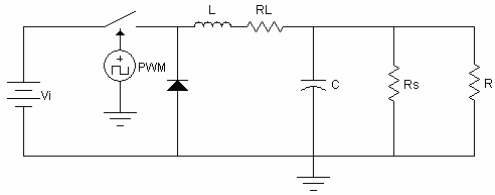


Fig 1. Buck converter.

Fig. 1 shows a Buck converter, where Vi is the input tension. The signal of commuted is given by a PWM generator, where L represents the coil that will store the levels of energy, given through the Vi. R_1 is the resistance of the losses of the coil. C is the capacity in which the tension signal is given to the coil filters. R_s is the drain resistance necessary to ensure a minimal discharge of the coil to avoid energy excess when the output resistance is elevated [8], and R corresponds to the charge resistance.

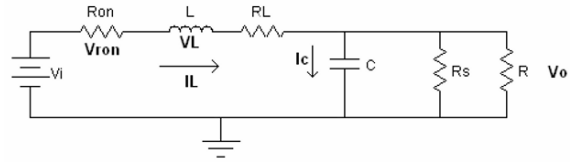


Fig 2. Buck converter for T_1.

See Fig 2. in T_1 the switch is close, in Vi conducts charge. L, R_{ON} corresponds to the loss of the dynamic resistance of the commuted device [8]. In this way it has:

$$v_L(t) = L * \frac{di_L}{dt} \quad (4)$$

$$i_C(t) = C * \frac{dv_c}{dt} \quad (5)$$

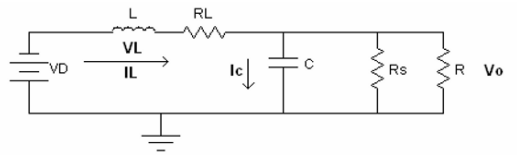
According to (5) this represents:

$$C * \frac{dv_0}{dt} = i_L - \frac{v_0}{R_{ol}} \quad (6)$$

Where:

$$R_{ol} = \frac{R * R_s}{R + R_s} \quad (7)$$

Fig 3. Buck converter for T_2



In Fig. 3, it is observed the circuit for T_2 in where the commuted device is open, not allowing the charge between Vi and L. At this moment the coil starts its discharge through the diode, where V_d corresponds to the proper losses of the element according to [8]:

$$L * \frac{di_L}{dt} = -v_D - v_o \quad (8)$$

$$C * \frac{dv_0}{dt} = i_L - \frac{v_0}{R_{ol}} \quad (9)$$

The objective is to make a simulation of the current of Buck, it is necessary to build an average model between model T1 and model T2, it takes the current signal, obtaining the averages finally for T:

$$L * \frac{di_L}{dt} = \delta * v_i - \delta * i_L * (R_L + R_{ron}) - v_0 - (1 - \delta) * v_D \quad (10)$$

$$C * \frac{dv_0}{dt} = i_L - \frac{v_0}{R_{ol}} \quad (11)$$

Where R_{ol} is the converter charge, which is not constant, i_L is an approximation of the output current, because the current i_c is despicable compared with i_L.

3. Neural Networks And Inverse neural Control

The artificial neural networks are computational models trying to emulate the human brain behavior. They are formed by fundamental units called artificial neurons. Each neuron is composed by a certain number of inputs, and processed by known values called weights. Each input is related with a correspondent weight, afterwards these values are added. Where a Bias value is also added, this value provides to the sumatory of the input values, multiplied by their corresponding weights, a constant level; the total result of the sumatory, is processed by an activation function, this one could be a Non-linear type, by using Hyperbolic tangent sigmoid transfer function (Tansig)[10], in the hidden layer, and Linear transfer

function (Purelin)[10], the multilayer perceptron is a universal approximator [13], it is a necessary condition for the identification of dynamical systems. This model is known as perceptron, is presented in Fig. 4.

The multilayer perceptron was developed, because the perceptron does not solve linear separable functions therefore, it was necessary to develop multilayer perceptron. This can solve the trouble of the linear separability and allows the neural networks to learn more complex functions, and at the time, dynamic systems [9].

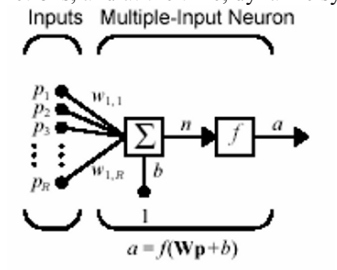


Fig 4. [10] Perceptron.

The neural networks are trained through input and output patterns, in these cases the neural network is started with random weights and Bias. Afterwards, the input vector of the neural network is given, the output value is measured, the main square error (MSE) is found between the desired output for the corresponding input, and the obtained output:

$$E(k) = \frac{1}{2} * \sum_{j=1}^{N_E} (y_d - y_N)^2 \tag{12}$$

Where N_E corresponds to the number of input-output patterns, y_d is the desired output and y_N the obtained output value. Later, with this error value, the gradient by descent technique and the Back Propagation algorithm are used, to refresh weights and Bias values:

$$w^R(t+1) = w^R(t) - \alpha \Delta w^R \tag{13}$$

Where α corresponds to the training factor, and:

$$\Delta w^R = -\frac{\partial E}{\partial w^R} = (y_d - y_N) * f'(n) * p_R \tag{14}$$

That corresponds to gradient [10].

In the neural control, the objective of the training for the Neural Network is to supply the desired behavior in the signal of the neural controller; to search the desired signal in the output of the system to control; the MSE is minimized, between the desired output response controller and the obtained output response controller, with an input-output pattern. It is necessary to use a regression vector in the input and the output patterns for the training, in order to identify the dynamic of the system [11].

The Inverse Neural Controller wants to generalize the inverse dynamic of the system to control through a neural network; this can be seen in Fig. 5.

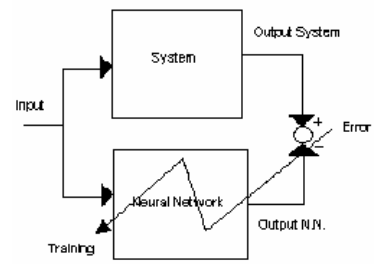


Fig. 5. Inverse Neural Controller.

Assuming that the system to control can be described as:

$$y(t+1) = g[y(t),...,y(t-n+1),u(t),...,u(t-m)] \tag{15}$$

The output desired for the inverse controller can be estimated by:

$$\hat{u}(t) = \hat{g}^{-1}[y(t+1), y(t),..., y(t-n+1), u(t),...,u(t-m)] \tag{16}$$

Where the signal *y(t+1), y(t),...,y(t-n+1)* is replaced by the desired output.

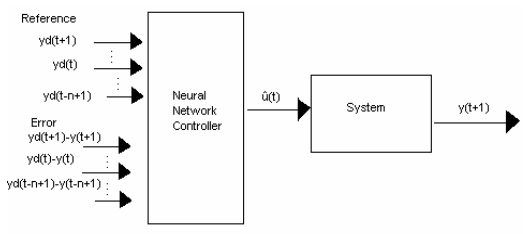
The NNARMAX (Neural Network AutoRegressive Moving Average eXternal input) model structure used in this paper for the Inverse neural controller, is showed in Fig. 6, the regressor yd*(t+1), $y_d(t)...y_d(t-n+1)$,* is taken as the reference. The error regressor is taken between the reference y_d and the output y.

Then:

$$\varphi(t,w) = [y_d(t+1),..., y_d(t), (y_d(t+1) - y(t+1)),..., (y_d(t) - y(t)]^T \tag{17}$$

and:

$$\hat{u}(t,w) = f(\varphi(t,w), w+B) \tag{18}$$



Where û(t,w) is the output of the neural controller, φ(t,w), w + B is the input of the neural network multiplied by the weights, B the bias of the neural output, and the f() is the function of activation.

Fig 6. NNARMAX model for inverse neural controller.

5. Simulations and Results

Using (10) and (11), the buck converter is modeled, for the simulations using Simulink of Matlab, in front of this design the next values are obtained [8]:

$R_{ON} = 0.5\Omega$
L = 0.001 H
Vi = 24 V.
Vd = 0.7V.
C = 1 mF.
$R_s = 24\ \Omega$

Output resistance is variable between 0Ω and 1.5 Ω, and δ takes values as $0 \leq \delta \leq 1$. Two layers perceptron will be used, the neurons in the hidden layers will change according to each case, one neuron in the out layer; the sample time for all cases of simulation are 0.1 time units, functions Tansig and Purelin respectively, the signals of voltage and current have 4 delays in the input of the neural control, and error signal with 3 delays [7], using general inverse neural control.

Case 1: The neural control was trained with 6 neurons in the hidden layer; using the structure NNARMAX with current reference of 8 A; the signal error is taken between the reference and the signal current of the Buck, the number of input – output patrons for the training was 60000 patrons (6000 time units per 0.1 time units for sample time), like the conditions of the current control model are rough and the system is strongly nonlinear with hard charge [8], 40000 patrons with randomly charges between 0Ω and 0.5Ω were taken, and 20000 patrons between 0.5Ω and 1.5Ω (nominal charge), this way it allows specialized training in hard conditions of the system to control [7] , with 25 epoch of training, the MSE converged to 0.0402585; for the validation, the neural control are simulated with randomly charges varying between 0Ω and 1.5 Ω every 1 time unit, the result of the simulation is presented in Fig. 7.

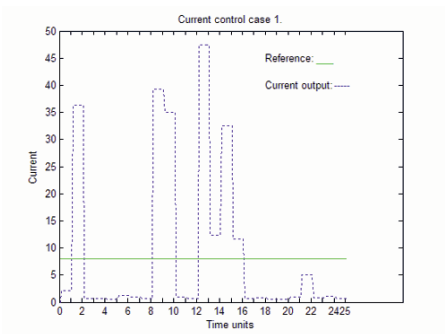


Fig. 7. Current case 1

The MSE (Squart Mean Error) developed during the simulation is presented in Fig. 8.

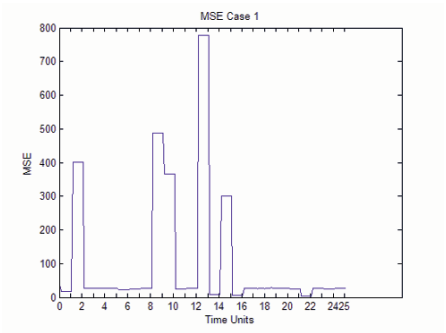


Fig 8. MSE case 1

The MSE is strongly hard, controller with those characteristics is not viable.

Case 2: Waiting for a better response, the number of neurons in the hidden layer are increased to 12, de others parameters are equal, in 50 epochs the MSE converged to 0.03946, the characteristics of the simulations for validation are the same that in case 1, the output current is showed in Fig.9 and the MSE in Fig.10.

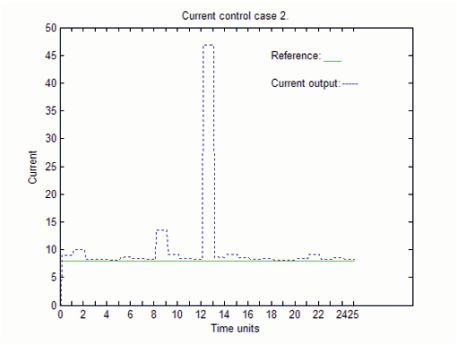


Fig. 9. Current case 2

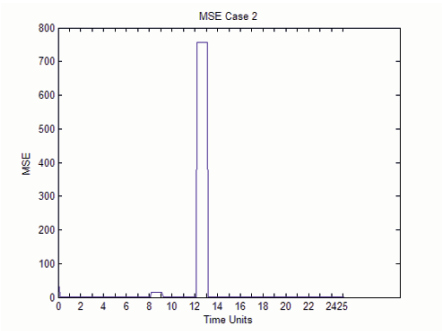


Fig.10. MSE case 2.

The MSE is less than case 1, but there is not short circuit protection, in times 12 and 13, there is a strongly charge, tendent to short circuit; although increasing the number of neurons in the hidden layer improves the controller, it does not seem to improve the short circuit protection.

Case 3: Looking for a short circuit response, the purpose is to add input-output patrons to 120000 patrons, 12000 time units per 0.1 time units like sample time), 40000 are with charges between 0Ω and 0.1Ω, 40000 patrons with charges between 0.1Ω and 1Ω, 20000 patrons between 1Ω and 1.5Ω, and other 20000 patrons with all range between 0Ω and 1.5Ω, with 5 neurons in the hidden layers, other parameters are the same that in the last cases; with 25 epochs of training, the MSE was 0.03142; for the validation, when it is used the same characteristics of the previous cases, presents a dc level in the current output, a solution is to use a precompensator , in the input reference current signal, and signal of error (cascade configuration between reference and comparator), the advantage of this configuration is that it doesn't affect the roots and the stability of the system [12] (remember there is not a mathematical model to estimate the stability), and its possible to tune the precompensator with out altering the stability.

Using trial and error, the value found for the precompensator for the reference is 0.1667, the compensator goes to reference input and error input, the signal current is presented in Fig.11, the MSE in Fig.12.
A better following of the signal of reference exists, and there are some control responses of short circuit in times 12 and 13, but the different values of MSE establishes that the controller is not efficient for the different values of charge.

Case 4: Finding a solution as proposed, is to divide in to modules the neural controller, modifying the precompensator, waiting a tuning in every range of charges, the tuning was done by trial and error. During the experimentation it was not enough one precompensator, it was necessary in every module a compensator for the reference current signal input and a compensator for reference for error signal input.

The choice of the modules was the charge ranges of the training, the first neural controller module for charges between 0Ω and 0.1Ω, the second module for charges between 0.1Ω and 1Ω, and the last module was for charges between 1Ω and 1.5Ω.

Module 1: The values found were: Current reference (K1): 0.68125, reference for signal error (K2): 0.375, during 25 epoch of training the MSE converged to 0.0314198

Module 2: The values found were: K1: 0.6625, K2: 0.375, during 25 epoch of training the MSE converged to 0.04026. Module 3: The values found were: K1: 0.4938, K2: 0.625, during 25 epoch of training the MSE converged to 0.04024.

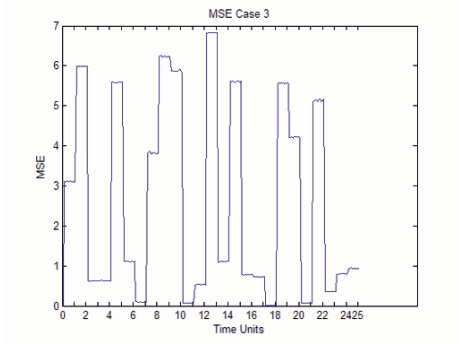


Fig. 11. Current case 3.

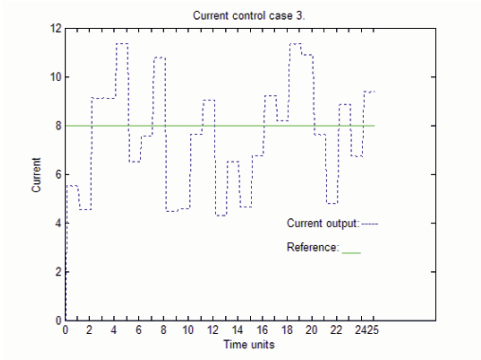


Fig. 12 MSE Case 3.

Any module independent is adequated inside its range, but the specification of the problem needs to control full range, then its necessary to build a selector of module control, for this particular simulation made in Matlab, an embedded Simulink Block is used, it selects the module depending on the sensed charge, the charge signal is taken to divide the voltage and the current. In the simulation, with the conditions of the previous presented cases, the current signal is in Fig.13, and MSE in Fig 14.

The MSE is less than other cases, presenting protection to short circuit, watch times 12 and 13, and there isn't any problem with the modular interaction, being this a viable proposal for a neural current control.

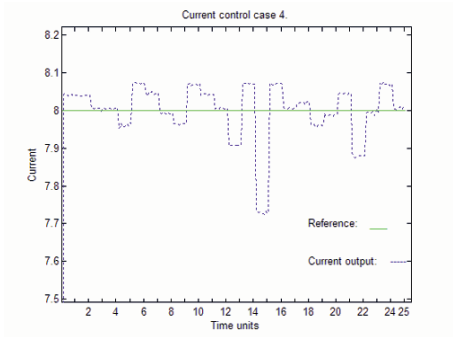


Fig. 13. Current Case 4. All modules.

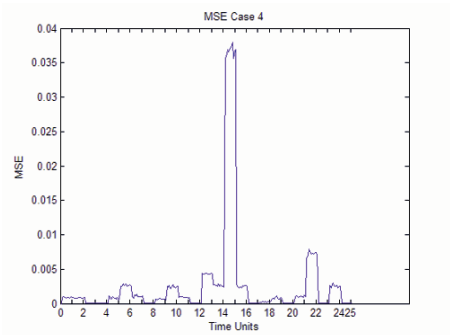


Fig. 14. MSE Case 4.

6. Conclusions and Future Works.

In its basic configuration, the inverse neural control is not an adequated controller for the current control in a Buck converter, specially searching for a current regulation in short circuit, when incrementing the input-output patrons for the training, and specializing the dates in charge regions, besides the simulation presents a stable behavior, with different states, depending on the value of the charge, including short circuit; when increasing the numbers of neurons in the hidden layer no better results were displayed as the ones described previously.

Upon this situation, it was proposed to train inverse neuronal controllers in certain regions of load of the output, these are in parallel and they are activated one by one, depending on the value of the tension and the current the corresponding controller is activated.

Making tests of trial and error, three modules were necessary, using precompensators in each module, persistent problems of error of stationary state were solved, without affecting the stability of the system. When making the simulations with this proposed system, the worse error had a smaller MSE of 0,04, including protection against short circuit. The simulations were made with Simulink of Matlab 7,0, using compatible elements with Real Time Toolbox, allowing in future works the study of this model for a future implementation in hardware. Another future work is to change the selector of modules, by a neuronal network that carries out itself work with similar behavior.

References

[1] H. Chan, K. Chau, C. Chan, "Methodology for Analysis of Neural Networks Generalization in Control Systems" , Department of Mechanical and Industrial Engineering, University of Toronto, 1997.

[2] I. Petrovic, A. Magzan, N. Peric, J.Matusko, "Neural Control of Boots Converters Input Current", Zagreb Croatia,2005.

[3] P. Isasi, I. Galván, *Redes Neuronales Artificiales Un Enfoque Práctico,* Prentice Hall. 2004.

[4] M. Norgaard, O. Ravn, N.K. Poulsen, L.K.Hansen, *Neural Networks for Modeling and Control of Dynamic Systems,* Springer, Technical University of Denmark.

[5] J. Mahadavi, N. Nasiri, A. Agah, A. Emadi. "Applications of neural Networks and State-Space Averaging to DC/DC Power Converters in Sliding Model Operation. ", IEEE Transaction on Mechatronics. Vol 10. No 1. 2005.

[6] P. Carbonell., J. Navarro. "Local Model Based Fuzzy Control of Switch Model DC/DC Converters", Proc. of the 14 th IFAC World Congress, 1999.

[7] J. Soriano, J. Guacaneme, J. Guarnizo, J. Lopez, "General Inverse Neural Control Nnarmax Structure For Buck Converter Using State Variable ", XVII Congreso Chileno de Control Automatico, 2007

[8] R. Ericcson, D. Maksimovic, *Fundamental of Power Electronics,* Kluwer Academic Publisher, University of Colorado, 2001

[9] S. Haykin, *Neural Networks A Comprehensive foundation,* Macmillan College Publishing Company, 1994.

[10] H. Demuth, M. Beale, *Neural network Toolbox For Use With Matlab,* 2004.

[11] M. Hagan, H. Demuth, O. De Jesus, "An Introduction to the Use of Neural Networks in Control Systems", Oklahoma State University, University of Colorado, 2005.

[12] B. Kuo. *Automatic Control Systems,* Seventh Edition, *Prentice* Hall, University of Illinois at Urbana-Champaign, 1996.

[13] Lo, T., -H, *"Multilayer Perceptron and Radial Basis functions are Universal Robust Approximators,"* Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on. Volume 2, Issue , 4-9 May 1998. Page(s):1311 - 1314 vol.2.

Management Study of Layered Architecture to Incorporate Mobile Devices and Grid Computing

Venkata Durga Kiran.Kasula, S.Srinivasa Rao, Shahana Bano
Lecturer, Department of C.S.E, Koneru Lakshmaiah College of Engineering, Green Fields, Guntur Dist, Andhra Pradesh
kvd_kiran@yahoo.com, srinu_mtech05@yahoo.co.in, tumpara@yahoo.co.in

Abstract **- This paper recommends a layered system model to overpass the gap between mobile and grid computing world. The model divides the complexities in mobile grid integration, among different components of different layers of the proposed model. The model proposes an efficient architecture which addresses the problem of scheduling and disconnection. The model addresses resource scheduling problem and principally focuses on disconnections problem of the mobile devices, battery power and economy issues.**

I. INTRODUCTION

The capability of running small computational tasks or jobs on demand makes it a computing element; a mobile device is able to register physical phenomena like sensors. It can provide intelligent user interface for remote analysis and control, and it can also provide persistent storage for data produced or configuration information used in the grid. The most natural way for small wireless devices is to provide ubiquitous access to timely information as data consumers. As such they provide handy alternative and extension to traditional access media such as standalone networked applications or Web browsers.

The chief restraint in Grid Computing technology, which is forcing the people to stay away from this technology and is not with in the reach of common people. By providing access to grid from a simple mobile device a common man can get benefit of this technology. So the integration of mobile and grid provides new opportunities, along with introduction of new challenges. Introduction of mobile devices in a distributed computing technology like grid compute requires an extension of the heterogeneity model. Model needs to consider the fact that grid may have much different kind of technologies which needs to be incorporated into grid nicely. Due to the mobility problem the network is unstable in wireless domain, disconnections and the resulting network partitions should be handled efficiently. Most of the projects propose solutions to disconnections by treating them as failures or willing permanent changes. However, with mobile or wearable devices, these events need to model in a general way. A mapping scheme should decide how different components of an application should be mapped, i.e., which components should be assigned to the back-end and which to the front-end (mobile device). These decisions should consider various issues including power consumption, locality of access, resource requirements, and throughput requirements.

The existing limitations in mobile devices make it beneficiary from the Grid. The main advantages of mobile Grid computing include mobile-to-mobile and mobile-to-desktop collaboration for resource sharing, improving user experience, convenience and contextual relevance and application scenarios. A grid-based mobile

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 123–127.

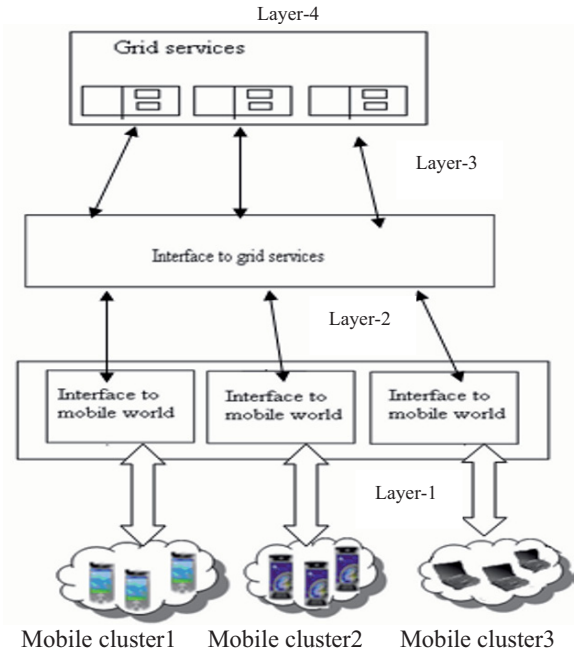


Figure 1 Proposed System Model

environment would allow mobile devices to become more efficient by offloading resource-demanding work to more powerful devices or computers.

Different people have implemented systems where the grid services are made accessible from mobile devices, all the existing implementations considered their own environment, but nobody have tried to give generic framework to integrate mobile and grid. Here system propose model, which addresses some issues in integration. The resource heterogeneity problem, resource scheduling problem, data representation problems, task decomposition issues are some of the issues which we considered here and also provide the details regarding the implementation of the proposed system model.

II. SYSTEM MODEL

The proposed system model is a Layered type (depicted in Figure 1). Due to the layered approach changes made in one layer not necessarily reflects in the other layers functionality and implementations, upper layer entities will provide services to the lower layer entities, and lower layer entities will depend on upper layer services. Due to layered approach different layers can be developed independently. With proper interface chord a mobile device can start communicating to the entities in the internet world appellation as Layer-1. The following model is introduced to address the requirements of various mobile technologies. These components should have features like encapsulating heterogeneity, able to communicate with standard protocols and languages like HTTP and XML, able to handle a lot many requests. In this context when mobile devices are acting as processing elements rather than just grid users, these entities required to have feature like dividing the work, assigning subtasks among the mobile devices, accumulating the processing results from different mobile devices before sending back the result named as Layer-2.

The enhancement in grid technologies will become independent of Layer-2 entities and acts as interface between grid services and mobile world interfaces. It takes the parameters from lower layers based on those parameters it will locate the appropriate services on the grid invoke the proper operation on that service recommend as Layer-3. This interface should also have the ability to invoke grid services which might have been implemented in various technologies. Layer-4 will have grid services which exploit idle resources spread across the internet. These services can be implemented using different technologies like condor, GT (Globus Tool kit) etc. Layer-4 entities will serve the requests passed from its lower layers and sends back the results. In these architecture mobile devices acting as grid elements and will send a task to the Layer-3 components. The Layer-3 would subsequently pass the task to Layer-2 component and sends back the processing results.

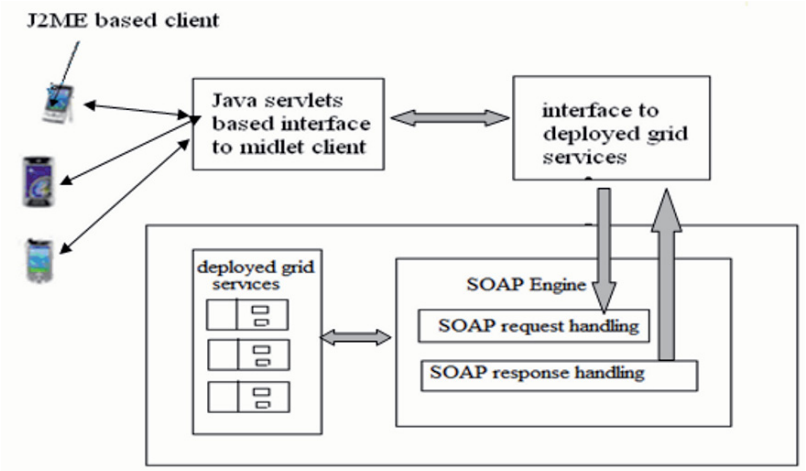


Figure 2 Architecture for Interface to GT based Grid Services through JAVA enabled mobile devices

III. PROPOSED SYSTEM MODEL

In the framework shown in Figure 2 the model depicted our attempt, to partial implementation of the System. The architecture for Interface can be broadly devised into five components. Different components will serve purpose of different layers entities. Though the following framework considers only JAVA enable mobiles, framework is easily extensible onto other technologies.

When a number of jobs arrive at the JQS (Job Queuing Server) of *mobile device interface*, or a number of workloads are decomposed from a single job, the JQS selects certain mobile nodes. In a static grid environment, we achieve improving performance as the number of participating computing nodes grows. However, the rule changes when the mobile devices participate in the computing. Links of some mobile nodes are unreliable, that is, they are likely to be in the disconnection state for a long time while others are not. The unreliable mobile nodes may decrease the overall job execution performance because the JQS should wait for the reconnection of any disconnected node in order to complete the jobs. We should determine whether it is better to include a mobile node in the computing or not. We might get different options of mobile devices sets, which reduce response time. We will try to increase the average battery power and also decrease the average cost of the selected devices

The framework concentrated on JAVA enabled devices dominated by JAVA most of the mobile device manufacturers are bringing JAVA supported devices into market. WSRF (Web Services resource framework) specifications are a core component in GT. It includes resource lifetime, resource properties, notifications, addressing specifications. The entire GT has been built on top of the WSRF component of GT. So by giving the interface to WSRF component through JAVA enabled mobiles, and can easily extend the interface to remaining components of the GT. The Java2 Platform, Micro Edition (J2ME) provides a robust, flexible environment for applications and TV set-top boxes, as well as a broad range of embedded devices. Here we are giving interface to WSRF component features of GT through mobile devices. The components description with implementation details of the framework is as follows.

A) J2ME midlet client:

This component serves the purpose of Layer-1 entities to some extent. It is a J2ME Midlet. It has the ability to provide nice GUI to JAVA enabled mobile devices. The mobile user will enter the data on the Midlet GUI which acts as parameters to the WSRF component functionality those data will be transmitted over the air to the appropriate servlet running on web server. In fact J2ME GUI feature used just to illustrate our implementation. Anyone who wants to use the framework can issue a request to this component to perform a particular task on the grid. Once the request issued the task invocation will be done automatically. It is basically a Midlet which is written in J2ME, it is able to create HTTP connections to an HttpServlet. This servlet sits in a HTTP web server (e.g. apache tomcat) and serves to HTTP requests. The following model had used WTK1.0.4_02 (Wireless Toolkit 1.0.4_02) emulator for testing our J2ME client component.

B) JAVA servlets based interface to Midlet client:

It consists of JAVA servlets. It will pack the parameters received from midlet client component and passes it to deployed grid services interface component. This module belongs to Layer-2 of the System model. This component is Java based that runs in a Web server. A servlet is a Java class and therefore needs to be executed in a Java VM by a service and call a servlet engine. The servlet engine loads the servlet class the first time that it is requests, or optionally already when the servlet engine is started. It then stays loaded to handle multiple requests until it is explicitly unloaded or the servlet engine is shut down.

C) Interface to Deployed grid services:

This component meets requirements of Layer-3. It will take parameters from midlet interface component using those it will invoke the appropriate methods in the specified grid service. It will return the results back to the midlet interface module. In case of error it will return the appropriate error code as a result. It is a JAVA based component. It also forwards the result of a task processed by mobile devices to some client in Internet. The framework took Globus toolkit based WSRF client software and modified the code according to current requirement and made this component to communicate both with midlet interface component and grid service component.

D) SOAP engine:

This is a piece of software which meets SOAP message processing issues and belongs to Layer-3. It knows how to handle SOAP requests and responses. A web service is deployed into an Axis message processing node using a deployment descriptor known as a Web Service Deployment Descriptor (WSDD). The WSDD describes how the various components installed in the Axis node are to be 'chained' together to process incoming and outgoing messages to the service. Models have used the apache tomcat SOAP processing capabilities. Apache Axis is a standard implementation of SOAP engine. This component can understand SOAP request/response messages.

E) Deployed grid services:

The following framework has developed these services using GT. It belongs to Layer-4 of the proposed model. These are some toy web services like counter service whose value can be accessed and modified by invoking operations like add, subtract, get value etc, a factory service which creates

service instance, a resource lifetime illustration service which illustrates features like resource destroy, scheduling resource termination time etc. These services basically illustrates WSRF specifications like WS-Resource properties, WS-Resource lifetime, WS-Base faults, WS-Resource addressing. We used WSDL, XML, GT, JAVA technologies in realization of grid services. They can exploit the available resources on the grid in an efficient, secured manner. It utilizes the GT inbuilt features like resource management, execution, and data management, fault detection and security.

CONCLUSION

Grid computing is the latest technology which is still under the research in academia. It has not gained enough significance in the enterprise world yet. In the direction of letting the common people access to grid service we have proposed a System Model. The framework gave the implementation details of system Model. In the implementation we considered JAVA based mobile devices in bottom layer, the grid services components at layer-4 has been built using GT. Our implemented framework is easily extensible onto other technologies. The paper primarily focused on issues like intermittent disconnection problem of mobile devices, battery power details, economy issues.

ACKNOWLEDGMENT

I am greatly delighted to place my most profound appreciation to K.Satyanarayana Chairman of K.L.C.E, Dr.L.S.S.Reddy Principal, S.Venkateswaralu Dean, S.Vasavi Head.Of.Dept,K.RajaSekhar,K.TirupathiRao P.Sai Kiran K.L.C.E under thier guidance and encouragement and kindness in giving me the opportunity to carry out the paper. Their pleasure nature, directions, concerns towards me and their readiness to share ideas enthused me and rejuvenated my efforts towards my goal.

REFERENCES

[1] F. Berman and R. Wolski. “The AppLes project: A status report,” Proceedings of the 8th NEC Research Symposium, Berlin, Germany, May 1997.

[2] J.Czyzyk, M.P.Mesnier, J. J. Mor´e, “The NEOS server”, IEEE Computational Science and Engineering,1998, vol. 5, page(s):68-75.

[3] S.Yi, M.Livny, “Extending the Condor Distributed System for Mobile Clients”, ACM Mobile computing and Communications Review, 1999, vol. 3(1), pp. 38-46.

[4] D. Abramson, J. Giddy, I. Foster, and L. Kotler. “High Performance Parametric Modeling with Nimrod/G: Killer Application for the Global Grid?” In Proceedings of the International Parallel and Distributed Processing Symposium, May 2000.

[5] Mauro, Maheswaran, Balasubramanian, Paul, Farag, “Mobile Interfaces to Computational Data and Service Grid Systems”, ACM SIGMOBILE Mobile Computing and Communications Review, Oct 2002, vol.6, Issue4.

[6] Sang-Min Park and Jai-Hoon Kim. "Chameleon: A Resource Scheduler in a Data Grid Environment," 2003 IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID'2003), Tokyo, Japan, May 2003.

[7] Miika Tuisku, “Wireless Java-enabled MIDP devices as peers in Grid infrastructure”, Annual cross grid project work shop &1st European acrossgrid conference, 2003.

[8] D. Chu, M. Humphrey, “Mobile OGSI.NET: Grid Computing on Mobile Devices”, 5th IEEE/ACM International Workshop on Grid Computing - GRID2004, Pittsburgh, PA, Nov 8 2004.

[9] [David E. Millard, Arouna Woukeu, Feng Tao, Hugh C. Davis, “Experiences with Writing Grid Clients for Mobile devices”. 1st International ELeGI Conference on Advanced Technology for Enhanced Learning, Vico Equense (Naples), Italy, March 2005.

Robust Control PID for Time Delays Systems

Laura E. Muñoz, Omar Santos, Virgilio López, Marco A. Paz*
Centro de Investigación en Tecnologías de Información y Sistemas, UAEH
Carretera Pachuca-Tulancingo Km. 4.5, Pachuca, Hgo. México
*ITESM, CCM. Calle del puente 222, Tlalpan, Mex. DF
ml123512@uaeh.reduaeh.mx, omarj@uaeh.reduaeh.mx, virgilio@uaeh.edu.mx, marco.paz@itesm.mx

***Abstract*—In this paper we present a robust stability analysis for a time-delay systems, in closed loop with proportional, proportional integral and proportional derivative (P, PI and PD respectively) controllers. The frequency approach is used in order to obtain necessary and sufficient conditions for the robust stability of the characteristic equation in closed loop and to tune the controller. The time domain analysis is used to obtain sufficient stability conditions for the system in closed loop with the PI controller, with nonlinear uncertainties. Also, the results are applied to a binary column distillation and coupled tanks, process commonly used in industry applications.**

I. INTRODUCTION

Time delay systems are frequently found in the industry, such as manufacturing system, turbojet engine, telecommunication, economic systems and chemical engineering systems. Therefore, the robust stability analysis for time delay systems has been widely achieved in the last decades [7]. For time delay systems when a simple control law is introduced, it may cause undesirable changes in the behavior of the system, such as instability and oscillations [7]. Besides, if the system has nonlinear uncertainties of non-model dynamics, then a chaotic behavior may occur [3]. Therefore, the study of stability regions is an interesting problem. On the other hand, the main goal in this contribution is to apply the analysis of robust stability to two systems widely used in industry: column distillation and coupled tanks.

Distillation is the most common unit operation in the chemical industry [8]. It is well-known that dynamic nonlinearities, control loop interactions, and system directionality can make the dual composition control in distillation columns a challenging problem. These issues can be effectively tackled by using advanced control systems, but in some cases the implementation of advanced controller are difficult to understand and expensive to maintain [1]. Then the PI controller is the most useful controller. However there does not exist an analysis of robust stability for a distillation plant. In the case of coupled tanks, these relate to fluid transport and liquid level control problems as they would typically occur in process control industries.

The paper is organized as follows: in Section II the problem statement is presented. The frequency domain analysis is presented in Section III. Robust stability using Linear Matrix Inequality (LMI) approach is given in Section IV. Section V introduces simulation results of the illustrative examples, the binary column distillation and the simulation results of two coupled tanks respectively. Concluding remarks are given in Section VI.

II. PROBLEM FORMULATION

Let us consider the following time delay system given in a transfer function form:

$$\frac{Y(s)}{U(s)} = \frac{Ke^{-sh}}{Ts+1} \tag{1}$$

When the input signal $U(s)$ is a PI controller, two gains must be determined: K_p and K_I. Some methods exist in order to determine these gains [5] but only a few works exist that study the robust stability analysis for parameters uncertainties or nonlinear uncertainties. Besides, since the PI controller has an integral part, the time domain analysis it is involved, then, a transformation on the space state representation it is introduced.

III. FREQUENCY DOMAIN ANALYSIS

In order to synthesize a PI controller for system (1) *D*-partitions method [4] is used. Considering that $U(s)$ is a proportional controller, the characteristic equation in closed loop is given by

$$Ts+1+K_pKe^{-sh}.$$

Now, according with the *D*-partitions method, the first boundary of the hyper surfaces is given when $s=0$

$$1+K_pK=0. \tag{2}$$

Therefore,

$$\underline{K_p} = \frac{-1}{K}, \forall K \neq 0\,. \tag{3}$$

Second boundary is given when $s=j\omega$

$$Tj\omega+1+K_pK(\cos\omega h - j\sin\omega h)=0.$$

It follows that,

$$1+K_pK\cos\omega h = 0 \tag{4}$$

$$T\omega - K_pK\sin\omega h = 0\,. \tag{5}$$

After basic operations we obtain

$$\overline{K_p} = \frac{\sqrt{1+T^2\omega^2}}{K^2}, \tag{6}$$

where ω is the numeric solution of (4) and (5). Thus, (2) will be stable for any $K_p \in [\overline{K_p}, \underline{K_p}]$.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 128–133.

Now, we find the stability zone for K_I with some $K_p \in [\overline{K_p}, \underline{K_p}]$ fixed. The characteristic equation in closed loop using a PI controller is

$$Ts^2 + \left(1 + K_p K e^{-sh}\right)s + K_I K e^{-sh}. \tag{7}$$

The first boundary ($s = 0$) is

$$K_I = 0, \forall K \neq 0. \tag{8}$$

The second boundary ($s = j\omega$) for K_I is given by

$$\begin{aligned} -T\omega^2 + KK_I \cos\omega h + KK_p \omega \sin\omega h = 0 \\ -\omega + KK_I \sin\omega h - KK_p \omega \sin\omega h = 0.. \end{aligned} \tag{9}$$

Straightforward operations give

$$\overline{K_I} = \frac{\omega\sqrt{\omega^2 T^2 + 1 - K_p{}^2 K^2}}{K}. \tag{10}$$

Where ω is the numeric solution of (9). Therefore if $K_I \in [0, \overline{K_I}]$ and $K_p \in [\overline{K_p}, \underline{K_p}]$, then the characteristic equation given by (7) will be stable. Now we analyze the robust stability for the parameters K and h of the system in closed loop using the proportional controller. Consider the equations given in (4) and (5), making some operations, we arrive to

$$\omega = \sqrt{\frac{K_p{}^2 K^2 - 1}{T^2}}. \tag{11}$$

Rewriting (4) and (5) as

$$K_p K = \frac{\omega T}{\sin\omega h}, \; K_p K \frac{-1}{\cos\omega h},$$

It follows that

$$h = \frac{\tan^{-1}(T\omega)}{\omega}, \; \omega \neq 0. \tag{12}$$

Finally, replacing (11) in (12) we obtain

$$h = \frac{\tan^{-1}\left(\sqrt{K_p{}^2 K^2 - 1}\right)T}{\sqrt{K_p{}^2 K^2 - 1}}. \tag{13}$$

Remark 1: It is possible to find different zones of robust stability for the system (1) depending on the gain K_p values in the controller. Moreover, if the system has multiples controllers, we can obtain the stable values for K_p and K_I by tuning one by one, and finding the intersections of the values of K_p and K_I.

IV. TIME DOMAIN ANALYSIS

In this section the robust stability analysis of a time delays system with nonlinear uncertainties and with a PI controller is introduced. From system (1) it is feasible to obtain the following state-space representation

$$\begin{aligned} \dot{x}(t) &= Ax(t) + \sum_{i=1}^{m} B_i \bar{u}(t - h_i) \\ y(t) &= C_1 x(t) + C_2 d(t) \end{aligned} \tag{14}$$

where $x(t) \in \mathrm{R}^{\mathrm{n}}$, $A \in \mathrm{R}^{\mathrm{n\times n}}$, and $B_i \in \mathrm{R}^{\mathrm{n\times q}}$, $\bar{u}(t) \in \mathrm{R}^{\mathrm{q}}$ is the input control, $y(t) \in \mathrm{R}^{\mathrm{q}}$ is the output, $C_1(t) \in \mathrm{R}^{\mathrm{q\times n}}$, $C_2(t) \in \mathrm{R}^{\mathrm{q\times n}}$, $d(t) \in \mathrm{R}^{\mathrm{n}}$ is the vector of disturbance and m is the number of time delays. Notice that,

$$\begin{aligned} \bar{u}(t) &= \begin{bmatrix} u_1(t) & u_2(t) & \mathrm{K} & u_q(t) \end{bmatrix}^T \\ \bar{r}(t) &= \begin{bmatrix} r_1(t) & r_2(t) & \mathrm{K} & r_q(t) \end{bmatrix}^T, \end{aligned}$$

where,

$$u_k(t) = K_{pk} e_k(t) + K_{Ik} \int_0^t e_k(t)dt, \; k = 1,\ldots,q \tag{15}$$

$$e_k(t) = r_k(t) - y_k(t), \; k = 1,\ldots,q$$

and rewriting (15) in a matricial form, it follows

$$\bar{u}(t) = diag\{K_{pk}\}\bar{e}(t) + diag\{K_{Ik}\}\int_0^t \bar{e}(t)dt$$

where,

$$\bar{e}(t) = \begin{bmatrix} e_1(t) & \mathrm{K} & e_q(t) \end{bmatrix}^T.$$

If we choose

$$x_r(t) = \bar{e}(t)$$

then

$$\bar{u}(t) = \overline{K_p}\left(\bar{r}(t) - y(t)\right) + \overline{K}_I x_r(t)$$

Thus, (14) can be rewriting as:

$$\dot{x}(t) = Ax(t) + \sum_{i=1}^{m}\left(B_i\left(\overline{K_p}\bar{r}(t-h_i) - \overline{K_p}C_1 x(t-h_i) - \overline{K_p}C_2 d(t-h_i) + \overline{K_I}x_r(t-h_i)\right)\right)$$

$$\dot{x}_r(t) = \bar{r}(t) - y(t) = \bar{r}(t) - C_1 x(t) - C_2 d(t). \tag{16}$$

Without loss of generality, it is assumed that system (14) has an equilibrium point at the origin. When $r = 0$, and by considering that h_i is uncertain, the system has time varying delays bounded by

$$0 \leq h_i(t) \leq \bar{h} \quad \dot{h}_i(t) \leq h_d < 1$$

where $\bar{h}$ and h_d are constants. Then, (16) is rewritten as:

$$\dot{\hat{x}} = \sum_{i=0}^{m}\left(\hat{A}_i \hat{x}(t - h_i(t))\right) + \sum_{i=0}^{m}\left(f_i d(t - h_i(t))\right) \tag{17}$$

where

$$\hat{x}(t) = \begin{bmatrix} x(t) \\ x_r(t) \end{bmatrix}, \; \hat{A}_0 = \begin{bmatrix} A_{0(n\times n)} & 0_{(n\times p)} \\ C_{1(p\times n)} & 0_{(p\times p)} \end{bmatrix}$$

$$\hat{A}_1 = \begin{bmatrix} B_i\overline{K_p}C_{1(n\times n)} & B_i\overline{K_I}_{(n\times p)} \\ 0_{(p\times n)} & 0_{(p\times p)} \end{bmatrix}, i = 1,2,\ldots,m$$

$$f_0 = \begin{bmatrix} 0_{(n\times n)} \\ -C_{2(p\times n)} \end{bmatrix}, \; f_i = \begin{bmatrix} B_i\overline{K_p}C_{2(n\times n)} \\ 0_{(p\times n)} \end{bmatrix}, i = 1,2,\ldots,m$$

The initial condition of system (17) is given by

$$x(\theta) = \varphi(\theta), \forall\, \theta \in [-h, 0],$$

where $\varphi(\theta)$ is a piecewise continuous function. Assume that $\|f_i(d(t-h_i(t)))\| < M_i, i = 0,1,...,m$, represent the nonlinear perturbations of system (17) which satisfy

$$\|f_i(d(t-h_i(t)))\| \le \alpha_i \|f_i(d(t-h_i(t)))\| \tag{18}$$

where $\alpha_i > 1$ are given constants. Constraint (18) can be rewritten as

$$0 \le (\alpha_i^{\,2} - 1) f_i^T(d(t-h_i(t))) f_i(d(t-h_i(t))), i = 0,1,...,m \tag{19}$$

Using the LMI approach, a Lyapunov Krasovskii functional and the S-procedure [11], for the inequalities given in (19), sufficient conditions for robust stability are found. The sufficient robust stability conditions of system (17), with uncertainties (18), are given in the following delay-dependent stability result.

Theorem 2: The system described by (17), with uncertainties described by (18) is asymptotically stable if there exists a real matrix W, symmetric positive matrices $P, Q_i, V (i = 1,...,m)$ and scalars $\varepsilon_n \ge 0\ (n = 0,...,m)$ such that the following LMI holds:[1]

$$\begin{bmatrix}
R_{(1,1)} & K_{(1,2)} & K_{(1,3)} & \cdots & K_{(1,j)} & P_{(1,1)} & P_{(1,2)} \\
(*) & J_{(1,1)} & 0 & \cdots & 0 & 0 & 0 \\
(*) & 0 & J_{(2,2)} & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
(*) & 0 & 0 & \cdots & J_{(3,3)} & 0 & 0 \\
(*) & 0 & 0 & \cdots & 0 & M_{(1,1)} & 0 \\
(*) & 0 & 0 & \cdots & 0 & 0 & M_{(2,2)} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
(*) & 0 & 0 & \cdots & 0 & 0 & 0 \\
(*) & (*) & (*) & \cdots & (*) & (*) & (*) \\
(*) & (*) & (*) & \cdots & (*) & (*) & (*) \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
(*) & (*) & (*) & \cdots & (*) & (*) & (*) \\
(*) & 0 & 0 & \cdots & 0 & 0 & 0
\end{bmatrix}$$

$$\begin{bmatrix}
\cdots & P_{(1,j)} & G_{(1,1)} & G_{(1,2)} & \cdots & G_{(1,j)} & L_{(1,2)} \\
\cdots & 0 & G_{(2,1)} & G_{(2,2)} & \cdots & G_{(2,j)} & 0 \\
\cdots & 0 & G_{(3,1)} & G_{(3,2)} & \cdots & G_{(3,j)} & 0 \\
\ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\cdots & 0 & G_{(n,1)} & G_{(n,2)} & \cdots & G_{(n,j)} & 0 \\
\cdots & 0 & H_{(1,1)} & H_{(1,2)} & \cdots & H_{(1,j)} & 0 \\
\cdots & 0 & H_{(2,1)} & H_{(2,2)} & \cdots & H_{(2,j))} & 0 \\
\ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\cdots & M_{(3,3)} & H_{(n,1)} & H_{(n,2)} & \cdots & H_{(n,j)} & 0 \\
\cdots & (*) & -V & 0 & \cdots & 0 & 0 \\
\cdots & (*) & 0 & -V & \cdots & 0 & 0 \\
\ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\cdots & (*) & 0 & 0 & \cdots & -V & 0 \\
\cdots & 0 & 0 & 0 & \cdots & 0 & -V
\end{bmatrix} < 0 \tag{20}$$

where

$$R_{(1,1)} = \left(\hat{A}_0 + \sum_{i=1}^{m} \hat{A}_i\right)^T P + P\left(\hat{A}_0 + \sum_{i=1}^{m} \hat{A}_i\right) + \sum_{i=1}^{m} Q_i + W^T\left(\sum_{i=1}^{m} \hat{A}_i\right) + \left(\sum_{i=1}^{m} \hat{A}_i\right)^T W$$

$$K_{(1,j)} = -W^T \hat{A}_i,\ i = 1,...,m$$

$$P_{(1,j)} = P,\ i = 1,...,m+1$$

$$G_{(n,j)} = -\hat{A}_{n-1}^T \hat{A}_j^T V,\ n = 1,...,m+1,\ i = 1,...,m$$

$$L_{(1,1)} = m\bar{h}(W^T + P)$$

$$H_{(n,j)} = \hat{A}_j^T V,\ n = 1,...,m+1,\ j = 1,...,m$$

$$J_{(i,i)} = -Q_i(1-h_d),\ i = 1,...,m$$

$$M_{(n,n)} = -\varepsilon_n(\alpha_n^2 - 1)I,\ n = 0,...,m$$

In the next section the use of Theorem 2 is illustrated, in order to obtain the robust stability conditions for a binary column distillation.

V. ILUSTRATIVE EXAMPLES

In the following, some simulation results applied to a binary column distillation and to coupled tanks are presented.

A. Binary column distillation

Let us consider the following transfer function of a linear time delay system corresponding to a binary distillation column used for methanol-water separation given by Wood and Berry [10]

$$\bar{y}(s) = G(s)\bar{u}(s) + G_d(s)d(s) \tag{21}$$

where

$$G(s) = \begin{bmatrix} \dfrac{12.8e^{-s}}{16.7s+1} & \dfrac{-18.9e^{-3s}}{21.0s+1} \\ \dfrac{6.6e^{-7s}}{10.9s+1} & \dfrac{-19.4e^{-3s}}{14.4s+1} \end{bmatrix}, \tag{22}$$

$$G_d(s) = \begin{bmatrix} \dfrac{3.8e^{-8.1s}}{14.9s+1} \\ \dfrac{4.9e^{-3.4s}}{13.2s+1} \end{bmatrix} \tag{23}$$

and where the time constants and time delays are given in minutes and $y_1 =$ overhead mole fraction methanol $y_2 =$ bottoms mole fraction methanol, $u_1 =$ overhead reflux flow rate, $u_2 =$ bottoms steam flow rate, $d =$ column feed flow rate.

Consider (21), (22) and $d(s) = 0$ (this consideration is made only for the tune process). It is easy to see that,

$$y_1(s) = G_{11}(s)e^{-s}u_1(s) + G_{12}(s)e^{-3s}u_2(s)$$

$$y_2(s) = G_{21}(s)e^{-7s}u_1(s) + G_{22}(s)e^{-3s}u_2(s)$$

As u_1 and u_2 are PI controllers, it is necessary to obtain the gains for the PI controller. Following the approach shown in section III, it is necessary to obtain the values for K_p and then

[1] The proof. is shown in the appendix

the values for K_I.

By making $u_2(t)=0$, it has the characteristic equation given by (2). Using equations given by (3) and (6), it can be found the values for K_p for $G_{11}(s)$ and $G_{21}(s)$. These values are

$$K_{p11}=[-0.0781,\ 2.0994] \qquad K_{p21}=[-0.1515,\ 0.4726]$$

Therefore stable values for K_{p1} of $u_1(t)$ (associated to $G_{11}(s)$ and $G_{21}(s)$) are

$$K_{p1}\in[-0.0781,\ 0.4726].$$

Now, we find stable values for K_{p2} by making $u_1(t)=0$ Using (3) and (6), we have that

$$K_{p12}=[-0.6159,\ 0.0529] \qquad K_{p22}=[-0.4221,\ 0.0515]$$

Therefore stable values for K_{p2} of $u_2(t)$ (associated to $G_{12}(s)$ and $G_{22}(s)$) are

$$K_{p2}\in[-0.4221,\ 0.0515].$$

Once, we have the values for K_{p1} and K_{p2} the values for K_{I1} and K_{I2} can be found. The characteristic equation for system (21) is given by (7), thus we can apply (8) and (10) for each transfer function $G_{11}(s)$, $G_{12}(s)$, $G_{21}(s)$ and $G_{22}(s)$. Choosing $K_{p1}=0.2$ and $K_{p2}=-0.09$, we have that

$$K_{I11}=[0,\ 0.2627], \qquad K_{I12}=[-0.0419,\ 0],$$
$$K_{I21}=[0,\ 0.0367], \qquad K_{I22}=[-0.0425,\ 0].$$

Therefore the stable values for K_{I1} and K_{I2} are:

$$K_{I1}=[0,\ 0.0367], \qquad K_{I2}=[-0.0419,\ 0],$$

These values are used to analyze the robust stability for the binary column distillation. First, the robust stability for the proportional controller in closed loop using the *D*-partition method is analyzed. From (13), the stability zone for each transfer function $G_{11}(s)$, $G_{12}(s)$, $G_{21}(s)$ and $G_{22}(s)$ when have changes in T and h can be found. Results are shown in Figure 1.

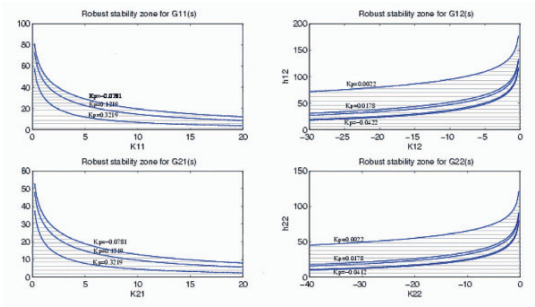


Fig. 1. Stability zones for $G(s)$

Now, consider the system given by (17), and the equation given by (22), for the column distillation $m=3$, and after some operations, the matrices $\hat{A}_i, i=0,1,...,3$, for the transformed system reads as:

$$\hat{A}_0=\begin{bmatrix}-0.6 & 0 & 0 & 0 & 0 & 0\\ 0 & -0.4762 & 0 & 0 & 0 & 0\\ 0 & 0 & -0.0917 & 0 & 0 & 0\\ 0 & 0 & 0 & -0.0694 & 0 & 0\\ 1 & -1 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & -1 & 0 & 0\end{bmatrix},$$

$$\hat{A}_1=\begin{bmatrix}-0.0168 & 0.0168 & 0 & 0 & -0.0054 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix},$$

$$\hat{A}_2=\begin{bmatrix}0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0.0810 & -0.0810 & 0 & 0.0017\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0.1212 & -0.1212 & 0 & 0.0026\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix},$$

$$\hat{A}_3=\begin{bmatrix}0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ -0.0133 & 0.0133 & 0 & 0 & -0.0042 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix}.$$

Thus, the robust stability for the column in closed loop using the PI controller, considering $\alpha_i=1.0001$, $i=0,1,...,3$ is achieved by using the LMI Toolbox of Matlab, it is found that the LMI given in Theorem 2 is feasible and the scalars are $\boldsymbol{\varepsilon_i=49.095, i=0,1,2,3}$ so the system is asymptotically stable under nonlinear disturbances.

B. Coupled tanks

Let us consider two coupled tanks, where the control problem is the regulation of the liquid level in one of the tanks by varying the speed of the circulating pump. After pulse testing the coupled tanks, the transfer function determined from data was

$$G(s)=\frac{Ke^{-hs}}{s}. \qquad (24)$$

Where $K=0.0097$ and $h=19\,\text{sec}$. According to section III, it is necessary to obtain the gains for the controller, in this case, instead of a PI controller, it has a PD controller. Then, first of all, the gain ..is obtained.

Then system characteristic equation with a proportional controller is

$$s + K_p K e^{-hs} = 0 .$$

According to the *D*-partitions method, the first boundary (for $s = 0$) is

$$K_p = 0 .$$

The second boundary is ($s = j\omega$)

$$j\omega + K_p K e^{-j\omega h} = 0 .$$

After some operations, second boundary follows

$$K_p = \sqrt{\frac{\omega^2}{K^2}} = \frac{\left|\omega^2\right|}{\left|K^2\right|}$$

By replacing the values for K and h; the stables values for K_p are

$$K_p \in [0, 8.5230].$$

Now the gain K_d can be found. The system characteristic equation in with a PD controller is given by

$$s\left(1 + KK_d e^{-hs}\right) + KK_p e^{-hs} = 0 .$$

According to the *D*-partitions method, the first boundary ($s = 0$) is

$$K_d = 0 .$$

The second boundary ($s = j\omega$) is given by

$$j\omega\left(1 + KK_d e^{-jwh}\right) + KK_p e^{-jwh} = 0 . \qquad (25)$$

After some operations, second boundary follows

$$K_d = \sqrt{\frac{\omega^2 - K^2 K_p^2}{K^2 \omega^2}} .$$

By choosing $K_p = 2$ the stable values for K_d are

$$K_d \in [0, 102.42] .$$

The robust stability zone is obtained using the (25) it follows that

$$\begin{aligned} \omega - KK_p \sin(\omega h) + KK_d \omega \cos(\omega h) &= 0 \\ KK_p \cos(\omega h) + KK_d \omega \sin(\omega h) &= 0 \end{aligned} \qquad (26)$$

Straightforward calculations give

$$\omega = \frac{KK_p}{\sqrt{1 - K^2 K_d^2}} . \qquad (27)$$

Equations (26) imply that

$$h = \frac{\cos^{-1}\left(-KK_d\right)}{\omega} . \qquad (28)$$

By replacing (27) in (28) it is possible to obtain the robust stability zone (Figure 2) of the parameters K and h for the model (24)

$$h = \frac{\cos^{-1}\left(-KK_d\right)\sqrt{1 - K^2 K_d^2}}{KK_p}$$

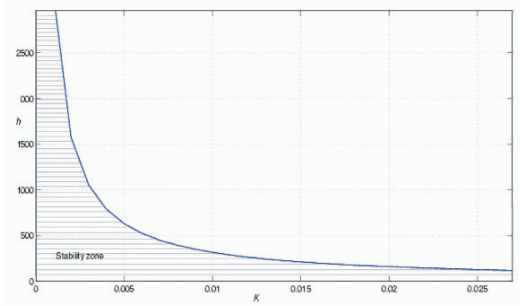


Fig. 2. Robust Stability Zone for $K_p = 2.$

VI. CONCLUSIONS

An analysis of robust stability for time delays systems was presented. This analysis is made with two different approaches: frequency domain using *D*-partitions method and time domain using Lyapunov-Krasovskii method and LMI. Also, the stability zones for gains of a PI and PD controller were obtained, and by using the frequency domain approach a stability analysis is introduced. Furthermore, for the time domain analysis of time delays system in closed loop with a PI controller, it was introduced a transformation of the state-space representation to replace the integral part of the controller, and in then the analysis using Lyapunov-Krasovskii LMI approach is easily realized. A similar procedure can be straightforward used when a PID controller is considered.

REFERENCES

[1] Bezzo F. et al. On the Controllability of Middle-Vessel Continuous Distillation Columns. *Ind. Eng. Chem. Res*., vol. 43, 2721-2729, 2004.
[2] Han, Q.L. and Yu, L. Robust stability of linear neutral systems with nonlinear parameter perturbations. *IEE Proc. Control Theory Appl*., vol. 151, no. 5, 539-546, 2004..
[3] Han des Heiden, U. et al. Existence of chaos in control systems with delayed feedback, *J. Diff. Eqs*., vol. 47, 273-295, 1983.
[4] Neimark, J. D-subdivisions and spaces of quasipolynomials. *Prik. Math. Mech*., vol., 13, 349-380, 1949.
[5] P. Cominos and N. Munro, PID controllers: recent tuning methods and design to specification. *IEEE Proc. Control Theory Appl*., vol. 149, 46-53, 2002.
[6] Park, P. A delay-dependent stability for class of linear systems with time-varying delay and nonlinear perturbations*, Comput. Math. Appl. An International Journal*, vol. 47, 1201-1209, 2004.
[7] Niculescu, S.I. *Delay Effects on stability: A robust control approach. Lectures notes in Control and Information Sciences*, 269, Springer Verlarg, Germany, 2001.
[8] Skogestad, S. Dynamics and control of distillation columns: a tutorial introduction. Institution of Chemical Engineers., *Trans. IChemE*, part A, 75, 539-562, 1997.
[9] Visioli, A. Optimal tuning of PID controllers for integral and unstable processes. *IEEE Proc. Control Theory Appl*. vol 143, 5, 429-435, 1996.
[10] Wood, R. K., and M. W. Berry, Terminal Composition Control of a Binary Column, *Chem. Eng. Sci.* 28, 1707, 1973 .
[11] Yakuvovich, V.A. S-procedure in nonlinear control theory. Vestn. Leningr. Univ. I. *Mat. Mekh*. 13, 62-77, 1971.

APPENDIX

Proof: [Theorem 2]. Choose a Lyapunov-Krasovskii functional candidate for system (17) as:

$$V(t) = V_1\left(x(t)\right) + V_2\left(t, x_t, \dot{x}(t)\right) + V_3\left(t, x_t\right),$$

where

$$V_1(x(t)) = x^T(t)Px(t)$$

$$V_2(t, x_t, \dot{x}_t) = \sum_{k=1}^{m} \int_{t-\bar{h}}^{t} (\bar{h} - t + \sigma)\dot{x}^T(\sigma)A_k^T X A_k \dot{x}(\sigma)d\sigma$$

$$V_3(t, x_t) = \sum_{k=1}^{m} \int_{t-h_i(t)}^{t} \dot{x}^T(\sigma)Q_i\, x(\sigma)d\sigma,$$

where symmetric positive definite matrices P, X, Q_i, are solutions of (20). The derivative of $V(t)$ along the trajectory of system (17) is given by

$$\begin{aligned}
\dot{V}_1(x(t)) &= 2x^T(t)P\dot{x}(t) \\
\dot{V}_2(t, x_t, \dot{x}_t) &= \sum_{k=1}^{m} \bar{h}\dot{x}(t)A_k^T X A_k \dot{x}(t) \\
&\quad - \sum_{k=1}^{m} \int_{t-\bar{h}}^{t} \dot{x}^T(\sigma)A_k^T X A_k \dot{x}(\sigma)d\sigma \\
\dot{V}_3(t, x_t) &= \sum_{k=1}^{m} x^T(t)Q_i x(t) \\
&\quad - x^T(t - h_i(t))Q_i x(t - h_i(t))(1 - h_d).
\end{aligned} \tag{29}$$

In order to obtain discrete-delay-dependent stability conditions, which includes the information of the time-delay $h(t)$, one usually uses the fact [2]

$$x(t - h_i(t)) \equiv x(t) - \int_{t-h_i(t)}^{t} \dot{x}(\sigma)d(\sigma), \tag{30}$$

to transform system (17) to the following one

$$\dot{x}(t) = \left(A_0 + \sum_{i=1}^{m} A_i\right)\dot{x}(t) - \sum_{i=1}^{m} A_i \int_{t-h_i(t)}^{t} \dot{x}(\sigma)d(\sigma) + \sum_{i=0}^{m} f_i(d(t - h_i(t)), t) \tag{31}$$

Replacing (31) in (29), we have

$$\begin{aligned}
\dot{V}_1(x(t)) &= 2x^T(t)P\left(A_0 + \sum_{i=1}^{m} A_i\right)\dot{x}(t) - 2x^T(t)P\times \\
&\sum_{i=1}^{m} A_i \int_{t-h_i(t)}^{t} \dot{x}(\sigma)d(\sigma) + 2x^T(t)P\sum_{i=0}^{m} f(d(t - h_i(t)), t)
\end{aligned}$$

Defining $b(\sigma) = Px(t), a(\sigma) = A_i\dot{x}(\sigma)$, using Lemma 1 in [6] and using the fact that $h_i \le \bar{h}, i = 1,..., m$, choosing $W = XMP$ and $V = m\bar{h}X$ we obtain that the derivative of $V(t)$ along to the trajectories of (17) is

$$\begin{aligned}
\dot{V}(t) \le\ & x^T(t)[(A_0 + \sum_{i=1}^{m} A_i^T)P + P(A_0 + \sum_{i=1}^{m} A_i) \\
& + (m\bar{h})^2(W^T + P)V^{-1}(W + P) + \sum_{i=1}^{m} Q_i]x(t) \\
& + x^T(t)W^T \sum_{i=1}^{m} A_i x(t) + x^T(t)(\sum_{i=1}^{m} A_i)^T W x(t) \\
& + [\sum_{i=0}^{m} (A_i x(t - h_i(t))) + \sum_{i=0}^{m} (f_i(d(t - h_i(t))))] \\
& \times \sum_{i=1}^{m} A_i^T V A_i [\sum_{i=0}^{m} (A_i x(t - h_i(t))) + \sum_{i=0}^{m} (f_i(d(t - h_i(t))))] \\
& + 2x^T(t)P\sum_{i=0}^{m} f_i(d(t - h_i(t), t))
\end{aligned} \tag{32}$$

Noting that using (17) it is possible that

$$\dot{V}(t) \le q^T(t)\xi q(t)$$

where $q(t) = \left(x^T(t)\ x^T(t - h_i(t))\ f_i^T\left(d(t - h_i(t)), t\right)\right)^T$ and

$$\xi = \begin{bmatrix}
K_{11} & K_{12} & \cdots & K_{1,p} \\
(*) & K_{2,2} & \cdots & K_{2,p} \\
\vdots & \vdots & \ddots & \vdots \\
(*) & (*) & \cdots & K_{m+1,m+1} \\
(*) & (*) & \cdots & (*) \\
(*) & (*) & \cdots & (*) \\
\vdots & (*) & \ddots & \vdots \\
(*) & (*) & \cdots & (*)
\end{bmatrix}$$

$$\begin{bmatrix}
J_{1,1} & J_{1,2} & \cdots & J_{1,m+1} \\
J_{2,1} & J_{2,2} & \cdots & J_{2,m+1} \\
\vdots & \vdots & \ddots & \vdots \\
J_{m+1,1} & J_{m+1,2} & \cdots & J_{m+1,m+1} \\
L_{1,1} & L_{1,2} & \cdots & L_{1,m+1} \\
L_{2,1} & L_{2,2} & \cdots & L_{2,m+1} \\
\vdots & \vdots & \ddots & \vdots \\
L_{m+1,1} & L_{m+1,2} & \cdots & L_{m+1,m+1}
\end{bmatrix}$$

Where

$$K_{11} = \left(A_0 + \sum_{i=1}^{m} A_i\right)^T P + P\left(A_0 + \sum_{i=1}^{m} A_i\right) + (m\bar{h})^2(W^T + P)V^{-1}(W + P)$$
$$+ \sum_{i=1}^{m} Q_i + W^T \sum_{i=1}^{m} A_i + \left(\sum_{i=1}^{m} A_i\right)^T W + A_0^T\left(\sum_{i=1}^{m} A_i^T V A_i\right)A_0$$

$$K_{1,p} = A_0^T\left(\sum_{i=1}^{m} A_i^T V A_i\right)A_{p-1} - W^T A_{p-1},\ p = 2,..., m+1$$

$$K_{q+1,q+1} = A_q^T\left(\sum_{i=1}^{m} A_i^T V A_i\right)A_q - Q_q(1 - h_d),\ q = 1,..., m$$

$$J_{1,p} = A_0^T\left(\sum_{i=1}^{m} A_i^T V A_i\right) + P,\ p = 1,..., m+1$$

$$J_{q+1,p+1} = A_q^T\left(\sum_{i=1}^{m} A_i^T V A_i\right),\ q = 1,..., m+1,\ p = 0,..., m+1$$

$$L_{q+1,p+1} = \left(\sum_{i=1}^{m} A_i^T V A_i\right),\ q = 0,..., m+1,\ p = 0,..., m+1$$

Using Schur complement we arrived to LMI given by (20). If there exist a real matrix W, symmetric positive de definite matrices P, Q_i and V such that,

$$\dot{V}(t) \le q^T(t)\xi q(t) < 0$$

and using the S-procedure [11] we can conclude this condition by finding non-negative scalars $\varepsilon_i \ge 0, i = 0,..., m$ such that

$$q^T(t)\xi q(t) + \sum_{i=0}^{m} \varepsilon_i\left(\alpha_i^2 - 1\right)f_i^T\left(d(t - h_i(t))\right)f_i\ \left(d(t - h_i(t))\right) < 0$$

for all $q^T(t) \ne 0$. Therefore, if exist a real matrix W symmetric positive definite matrices P, Q_i and V and the scalars $\varepsilon_i \ge 0, i = 0,..., m$ such that the LMI (20) holds, the system (17) with nonlinear uncertainties is asymptotically stable. ■■■

Wavelets vs Shape-Based Approaches for Image Indexing and Retrieval

L. Flores-Pulido[1], O. Starostenko[1], I. Kirschning[1], J.A. Chávez-Aragón[2], G. Burlak[3]
[1]Research Center CENTIA,Universidad de las Américas-Puebla, Mexico
aicitel_tryn@yahoo.com.mx, oleg.starostenko@udlap.mx, ingrid.kirschning@udlap.mx
[2]Universidad Autónoma de Tlaxcala, Apizaco, Mexico, achavez@ingenieria.uatx.mx
[3]Center for Research on Engineering and Applied Sciences, UAEM, Cuernavaca, Mexico, gburlak@uaem.mx

Abstract- **This paper presents a comparative analysis of some novel approaches proposed by authors for content based image retrieval (CBIR). One of them uses Two-Segments Turning Functions (2STF) and provides searching and retrieval of the multimedia documents within digital collections. Another technique retrieves images computing similarity between wavelet coefficients of querying and preprocessed images. For this purpose the Symlet transform has been used in designed system called Image Retrieval by Neural Network and Wavelet Coefficients RedNeW. However both of approaches operate with low-level characteristics processing color regions, shapes, texture, and they do not provide the analysis of image semantics. In order to improve these systems a novel approach is proposed that combines non-sensitive to spatial variations shape analysis of objects in image with their indexing by textual descriptions as part of semantic Web techniques. In the proposed approach the user's textual queries are converted to image features, which are used for images searching, indexing, interpretation, and retrieval. A decision about similarity between retrieved images and user's query is taken computing the shape convergence and matching to ontological annotations of objects in image providing in this way definition of the machine-understandable semantics. In order to evaluate the proposed approach the Image Retrieval by Ontological Description of Shapes IRONS system has been designed and tested using some standard domains of images.**

***Index Terms*— CBIR systems, ontological descriptions, shape indexing, visual information retrieval, wavelets**

I. Introduction

In well-known content-based image retrieval systems the extraction of image features is the principal procedure used for visual information indexing and interpretation [1], [2]. There are a lot of reports about novel approaches and methods for searching, classification, and retrieval of visual information on base of analysis of low-level image features, such as a color, texture, shape, etc. [3]. These methods sometimes are slow and too complex for design of real-time applications. Moreover, these CBIR systems do not define understandable semantics associated with automatic indexing and interpretation of visual information. Another significant problem is computing similarity between feature vectors representing images without their adjustment and normalization [5].

Among commercial CBIR systems that may be used as prototype for development of novel image retrieval techniques, there is one called CIRES (Content Based Image REtrieval System) that provides features retrieval, such as a structure, color, and texture of image combining them with user specifications of importance of these features in a query [2]. SIMPLIcity (Semantics-sensitive Integrated Matching for Picture Libraries) achieves image retrieval from the Web using texture, indexing by clustering of image segments, and feature vectors is generated by wavelet transform [6]. Another CBIR facility combining neural networks and matching wavelet characteristics is known as RedNeW system [7]. This one provides a region growing using multiresolution in YIQ color domain and applies Jacobs' metrics for computing similarity between retrieved images and visual query. The color and shape features are exploited in the system that is based on Two-Segments Turning function 2STF. Using the tangent space domain the invariance to rotation and scaling has been achieved as one of important and useful properties of the shape indexing process [3].

After testing these systems some disadvantages have been detected, such as a low percentage of relevant retrieved images and significant time for computing similarity between feature vectors. Hence, we propose to apply a novel approach for design of CBIR systems using the both an image semantics and its low-level features. The possible applications of the proposed image retrieval approach are the systems for supporting digital image processing services, high performance exchange of multimedia data in distributed collaborative and learning environments, medical and biological researches, digital libraries, and others where information is presented in visual form.

II. Related Approaches and RedNeW System

Among the image features the most important are color, shape, and region because they define specific regions in image, which may be interpreted and classified as objects with certain significance in a scene. The main purpose of image retrieval system is its content extraction that permits to choose the relevant information to user's visual query as it done now in well-known searching engine for textual documents. The well-known methods for global color/shape/region descriptions, such as Elasticity correspondence [8], Curvature scale space approaches [9], B-splines and chain case codes [5], Two-Segment Turning function [1], Fourier spectral descriptors, wavelets [10], etc. sometimes are too complex for

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 134–139.

fast non-sensitive to spatial variations processing and frequently they do not provide the efficient feature extraction. Anyway, some of these methods have been selected taking into account their performance, usefulness for improving the CBIR process, processing speed, simplicity to compute grade of similarity to input visual query, low number of iteration in retrieval process, and efficiency for description of image semantics.

The selected CBIR facility, which may be improved applying a novel retrieval approach, is RedNeW system presented in Fig. 1. It is based on wavelet transforms due to their high grade of relevance of retrieved images, satisfactory convergence of results, symmetry, and regularity useful for image processing with presence of noise [11]. The system designed by authors consists of two channels: one of them generates wavelets characteristics of the input querying image and another one retrieves the centroid corresponding to a group of images with similar meaning in preprocessed collection. The distribution of images in collection on base of their semantic features is achieved using the ART2 neural network. Neural net assigns to a set of image parameters corresponding values that define to which group this image belongs, for example, to group of buildings, maps, animals, landscapes, etc.

For each classified image in a group (in-class images) wavelet coefficients are computed, and then they are used for generation of centroid by calculation of average value of wavelet coefficients for particular class. However, these centroids do not express the general means of the images in class, they show how similar is one image to other in collection taking into account only low-level image features. That is why, the user feedback is needed to define the semantics of each image with user-oriented indexing vocabulary. An input querying image is divided in regions applying YIQ multiresolution decomposition and then sixteen characteristics that compose the feature vector, such as region area, its maximum radius and luminosity, RGB average, elliptical envelope, moments, etc. are computed for each region. The highest grade of similarity between this vector and centroids defines the group of images that may be candidates for retrieval. This grade is computed according the equation proposed by Jacobs [11].

$$\|Q,T\| = w_{(0.0)}\left|Q[0,0]-T[0,0]\right| + \sum_{j,j} w_{(i.j)}\left|\widetilde{Q}[i,j]-\widetilde{T}[i,j]\right| \quad (1)$$

where Q represent the feature vector of a querying image and T is the centroid corresponding to particular class, $w_{(i,j)}$ is a semantic weight associated with a class of images to which centroid belongs in T. When the best correspondence with one or more centroids is found the comparison of the wavelet characteristics of images within the selected class is applied to find those images that will have the highest grade of similarity. In this system Symlet transform with the best symmetry is used for computing wavelet coefficients shown in Fig. 2 cording to the following equation

$$Symlet_{(X)} = e^{\varphi(i)\omega} \quad (2)$$

where X is input image, $\varphi(i)$ is Daubechines transform for each pixel i, and ω is the moment of an image [7].

The disadvantage of RedNeW system is a significant time that it takes for organization of preprocessed collection that additionally requires the user's feedback description of images. We determined empirically that this time is about some seconds for retrieval of images from standard COIL_100 (Columbia Object Image Library) collection, which has 7200 images processed on personal computer of 2GHz and RAM of 1GB [12]. Additionally, if a new querying image has not the corresponding similar preprocessed images in collection, the approach is failed retrieving nonsense information. In this case the collection must be updated with a new image generating a particular semantic class and training of the neural network generating the corresponding centroid.

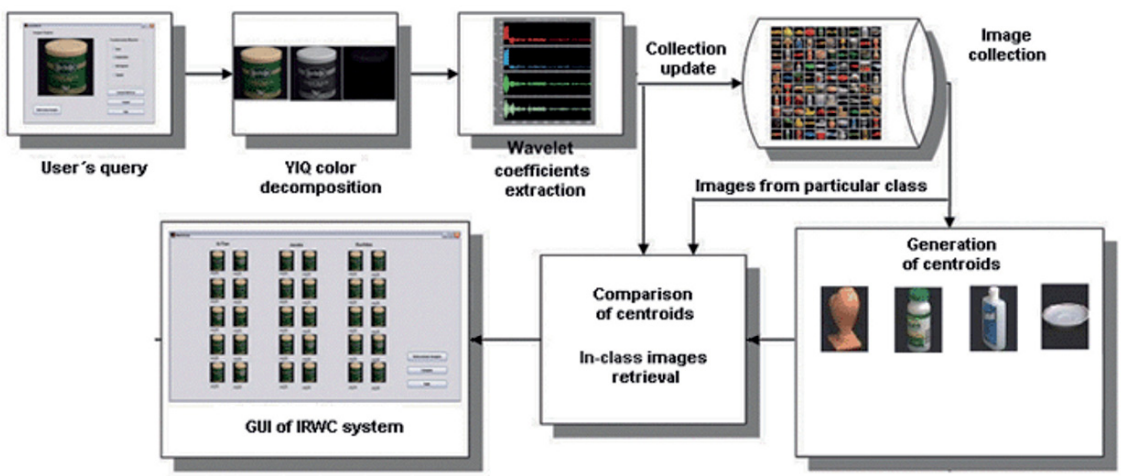


Fig. 1. Block diagram of RedNeW CBIR system.

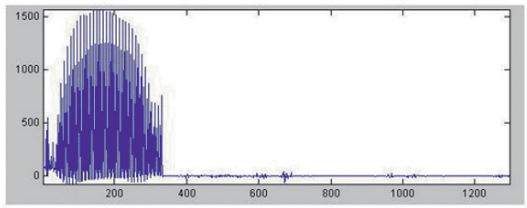


Fig. 2. Wavelet coefficients extraction by Symlet transforms

III. Shape-Based Retrieval Applying 2STF

Another well-known approach for CBIR exploits a shape as the more informational feature among low-level image characteristics. This system designed by authors is based on shape extraction using 2STF, the block diagram of which is shown Fig. 3 [1].

The frequent problem of shape representations is a great number of necessary vertices for polygon transformation, which must be reduced by proposed discrete curve evolution process**.** This technique reduces a set of vertices of a polygon to subset of vertices, which contain enough relevant information about the original outline by assigning a relevance measure to every vertex [13].

The image indexing based on shape is the main component of our proposed approach. We computed the similarity between two shapes applying the cumulative angle function or turning function. In general, the segments forming a polygon are represented by their length and absolute angle. In this way the shape is invariant to translation and scale, but it is sensitive to rotation.

To overcome this problem, we proposed an alternative way called two-segments turning function 2STF, where the relative angle between two consecutives edges of a polygon is computed instead of absolute one. That allows getting the same representation for a set of shapes even though they have been reflected or rotated. 2STF is calculated by traverse a digital curve in the counter clockwise direction assigning a certain value of 2STF for each line segment in a curve. In 2STF the steps on x - axis is defined by the normalized segment length, and the y - axis represents the turning angle between two consecutive segments. In Fig. 4 similar 2STFs is presented for polygonal curves with rotation and scaling. The similarity value between two shapes is based on analysis of matching their 2STFs.

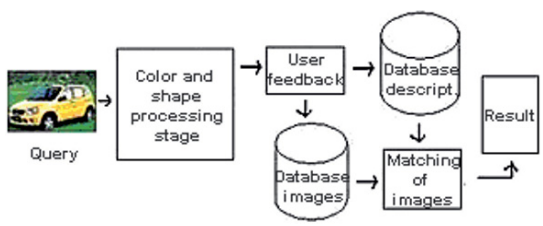


Fig. 3. Block diagram of the CBIR system based on 2STF

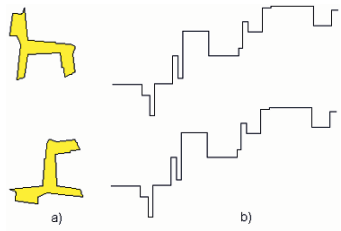


Fig. 4. Step 2STF functions for pattern in rotation and scaling

Basically, the proposed algorithm may be described as: the polygon simplified by curve evolution process is transposed into its 2STF; the resulting curves representing the polygon are scaled to the same length; one of the curves is shifted vertically over the second one to fit better; finally, the area between two curves is computed. The Fig. 5 shows the proposed matching strategy. The shaded area represents how similar these shapes are.

The disadvantage of 2STF is significant time that it takes to find the best correspondence between two curves. The experiments have been done on personal computer of 2GHz and RAM of 1GB taking into account valid combination of either concave or convex segments [1]. The time for matching of eight-segment polygons gives *0.375s* but for ten-segment shapes comparison the time is about *53.78s* that cannot be accepted as satisfactory result. This time may be reduced by decomposition of a polygon into groups G. This process consists in grouping the consecutive largest convex or concave segments covering the whole polygon. The number of operations for definition of groups is $N_G = \left(2^{M_a}\right)^2$, where M_a is the number of largest segments of a curve. Taking into account restrictions for grouping either concave or convex arcs the number of operations will be less, reducing at least in ten times the duration of matching process. Concluding, the advantage of this approach is independence from scale, reflection, translation, and rotation, but it requires a significant time for computing of similarity between shapes and the indexing process must pass through user's feedback annotations defining in this way the content of an image. These problems may be solved by another technique called Star Field representation of shapes.

IV. Design of Image Retrieval IRONS System

Proposed by authors Star Field (SF) is an alternative representation for shapes that allows to obtain more fast and precise comparison because it is not necessary to apply a great

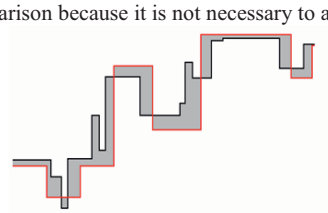


Fig. 5. Computing similarity between two polygons.

grade of evolution of polygonal curves. It means that we able to compare polygons with more than ten arcs (maximum value acceptable for 2STF). Our Star Field method combines approach for computing similarity among shapes proposed by Mokhtarian [9] and another one based on 2STF. He proposed to use the maxima of curvature zero-crossing contours of Curvature Scale Space (CSS) as a feature vector to represent shapes. However, computing CSS is expensive process, and we propose to use 2STF that is easier, faster and more effective [12].

Formally, a SF representation is a set of marks or stars $M_1;M_2::M_{nm}$, where *nm* is the number of vertices of the polygonal curve that it represents. This number is equivalent to the number of steps in its 2STF. A star is defined by means of two coordinates (*x, y*). The *x* - coordinates indicate the normalized distance from the starting point to the corresponding vertex, making sure that in the middle of the SF plane is the star that corresponds to the most important largest vertex of the polygon. The *y* - coordinate is the normalized angle between two consecutive segments that share the corresponding point. In other words, *y* - coordinate of stars correspond to the height of each step in its equivalent 2STF in the range of [0,1], where a value of zero represents a $-\pi$ angle and one corresponds to $+\pi$. The principal difference between 2STFs and SF is the grade of evolution of a curves. SF works with a larger number of vertices without increasing the time for their processing. The 2*D* plane of SF diagram is divided horizontally into two sections. The upper section holds the stars that represent vertices of concave segments. The lower part holds vertices of convex segments. In the Fig. 6 the apple-like 15-segment polygon and its Star field are shown.

For SF a new similarity measure using graph and adjacency matrix will be introduced. Given two polygonal curves *P1* and *P2* and their star field representations *SF1* and *SF2*, the graph *G* that allows us to compute their similarity is defined as $G = (V, E)$ where *V* and *E* are disjoint finite sets. We call *V* the vertex set and *E* the edge set of *G*. Our particular graph *G* has a set *V* that consists of two smaller subsets of vertices $v1$ and $v2$. Set $V = v1 \cup v2$, where $v1$ is the set of point of *SF1* and $v2$ is the set of points of *SF2*. The *E* is the set of pairs (r, s), where $r \in v1$ and $s \in v2$. Then we propose to use the adjacency matrix for representing the graph, where each cell of that matrix contains the cost for traveling from one column to each row and vice versa. The main idea behind the construction of the matching graph consists in building a connected weighted graph, so that an algorithm to find the minimum spanning tree is applied. The minimum spanning tree is a subset of edges, which forms a tree that includes every vertex, where the total weight of all edges in the tree is minimized. Thus, for the more similar shapes we obtain the lower value of corresponding total weight. SF permits to define a similarity measure based on calculation of a minimum spanning tree from a connected weighted graph, which is more fast and accurate. This new SF approach based on 2STF maintains its advantages and due to its simplicity allows working with complex polygons faster than using 2STF.

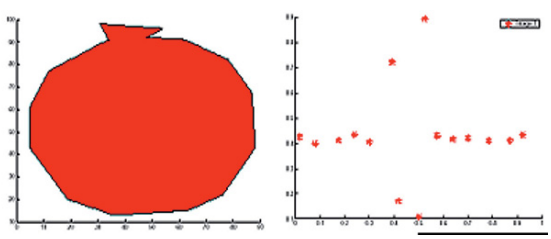


Fig. 6. The 15-segment polygon and its SF representation.

Additionally, in this approach we use hybrid feature vector, which defines low-level image characteristics (color/shape) and semantic descriptions. This permits to speed up the matching process as well as reduce the number of iterations with nonsense results. Second vector is formed by ontological description tool, which establishes the relationship between an object and its formal explicit definition. Hence, the meaning of an image may be obtained in textual form as a set of descriptions for each region related to the particular ontology. The Resource Description Framework (RDF) language to support the ontology management has been used in this approach that defines a syntactic convention and a simple data model to implement machine-readable semantics [14]. Using RDF it is possible to describe each web resource with relations to its object-attributes-value based on metadata standard developed by the World Wide Web Consortium [15]. The ontology is described by a directed acyclic graph. Each node has a feature vector that represents the concept associated with that node. Concept inclusion is represented by the IS-A inter-relationship. For example, particular elements of buildings, churches, etc. correspond to specific concepts of shapes defining these buildings, churches. If a query describes an object using this ontology, the system would recover shapes that contain windows, columns, façades, etc. even though, those images have not been labeled as geometric figures for the retrieved object. The feature vectors of each node in the ontology name space consist of keywords linking the previously classified images to the characteristics of a new shape extracted by the SF or 2STF. Finally, a decision about the images, which are candidates to be retrieved, is taken using Euclidean distance comparing two feature vectors as

$$d_E(\mu,\sigma) = \sqrt{\sum(\mu-\sigma)^2} \qquad (3)$$

where μ and σ denote low-level and ontological feature vectors for each image. The textual descriptions are found in ontological vocabulary and corresponding to these descriptions images in database are retrieved and shown in GUI of IRONS.

In order to evaluate the proposed approach, the Image Retrieval by Ontological description of Shapes (IRONS) system has been designed and tested. The block diagram of system is shown in Fig.6. Its input may be an image, shape, or a keyword. The retrieved images will be those, which have a high grade of similarity in color/shape and ontological annotations. The IRONS system consists of four principal modules: query preprocessing, indexing module, feature vector comparison, and feedback GUI.

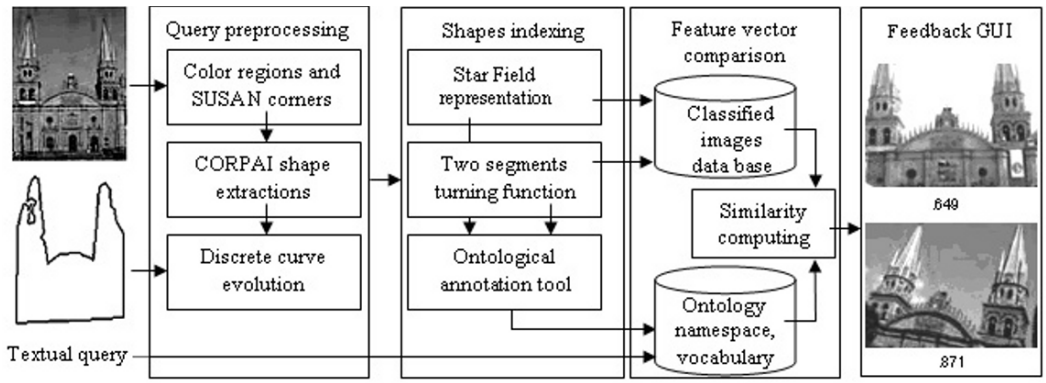


Fig. 6. IRONS system for shape indexing using the ontological descriptions

The query preprocessing module provides the selection of region containing the relevant objects. Once the sub-region is extracted, the object within that sub-region is found by the Convex Regions Preprocessing Algorithm in Images CORPAI and the Smallest Univalue Segment Assimilating Nucleus (SUSAN) algorithms [1]. The result of applying SUSAN and CORPAI algorithms is a convex polygon that may be simplified by discrete curve evolution process. The shape indexing module generates the low-level feature vector by SF or 2STF.

After comparison of this vector with others in classified image database the description of images with the highest grade of similarity are selected for visualization.

V. Experiments and Discussion

For estimating the performance of used approaches, particularly, wavelet, 2STF, and SF transforms, we divided experiments in two groups. First one is *Candidate Images Selected by Means of Low-Level Features experiment* performed over COIL_100 collection [16]. Each experiment consists of random selecting images form 72 different classes in collection of 7200 images. The purpose of this experiment is to observe how well the used techniques are able to choose the relevant images using only low-level image features. The second *Candidate Relevant Images Selected Using an Image Ontological Descriptions experiment* consists in evaluation of a retrieval process applying two feature vectors.

The evaluation of CBIR system is non-trivial task. This is because there is an amount of subjectivity involved into query interpretation by user. The similarity between a query and a set of retrieved images depends on individual perception of user. Nevertheless, there is a standard way of judging the obtained results. This technique consists in calculation of two metrics, such as recall and precision.

The recall measures the ability of a system to retrieve relevant information from all collection and may be computed as ratio *A/C*, where *A* is a set of relevant images retrieved by system, and *C* is a set of all relevant images in collection for particular query.

The precision metrics is the proportion of the number of relevant retrieved images to the total number of relevant ones in collection obtained from the ratio *A/B*, where *B* is a set of relevant and irrelevant images retrieved by system for a particular query. Table I shows the results of experiments with multiple queries and average recall and precision metrics for RedNeW, 2STF, and IRONS systems.

In Fig. 7 the retrieved without ontology images corresponding to a query “yellow circle” (first image in GUI) are presented. In Fig. 8 the retrieved images applying the ontology for query “blue vehicle” are shown.

TABLE I
The Reall/Precision for REDNEW, 2STF, and IRONS Systems

RedNew		2STF		IRONS shape only		IRONS with ontology	
Rec	Prec	Rec	Prec	Rec	Prec	Rec	Prec
0.25	0.53	0.32	0.6	0.38	0.6	0.7	0.88
0.15	0.42	0.27	0.65	0.6	0.7	0.6	0.92
0.05	0.16	0.32	0.4	0.55	0.45	0.72	0.95
0.2	0.45	0.32	0.45	0.6	0.67	0.7	0.96
16.3 %	39.0 %	30.8 %	52.5 %	53.2 %	60.5 %	68.0 %	92.8 %

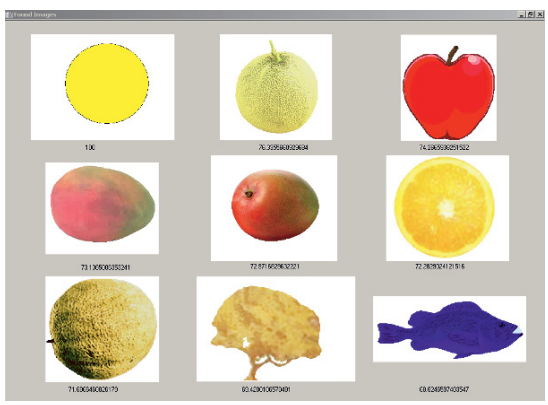


Fig. 7. Images retrieval without ontology.

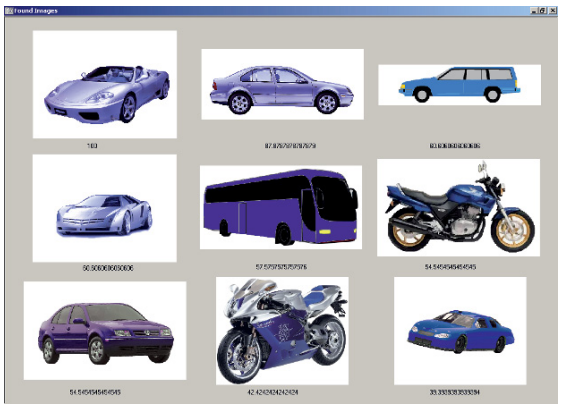


Fig. 8. Images retrieval with ontological description.

The difference between 2STF is that, the SF permits to process more complex polygons without their simplification. That is why, the increment of the recall/precision obtained by SF is due to more precise representation of shapes. The RedNeW and 2STF systems have been improved incrementing their performance for images retrieval and decreasing the number of iterations in searching process. The application of ontology allows us to reduce the semantic gap problem retrieving only relevant images from the same class to which the particular query belongs. As it shown in Table I, the precision with ontology achieves up to 92.8% of all relevant images.

It is possible to appreciate in Fig. 7 that the images have the similar color/shape features but belongs to different semantic classes. In the GUI the blue fish appears for yellow query. It is because the shape feature has as twice more relevance than color in system. It possible to appreciate in Fig. 8 that the IRONS system applying ontology retrieves images from the same semantic class and does not retrieve images from other ones, for example, fruits, buildings, animals, etc. reducing a set of irrelevant retrieved images and number of iteration needed to obtain satisfactory results.

VI. Conclusion

The evaluation of the proposed approaches and testing the designed RedNeW and 2STF systems show that ability of systems to retrieve relevant information from image collection is satisfactory. However, they have significant number of iterations and high probability to retrieve nonsense images.

In IRONS system satisfactory retrieval of expected images is provided faster due to the lower number of iterations as result of analysis only images, which belong to the same semantic class applying ontological annotations.

The most important contribution of this research is proposed hybrid approach combining the advantages of low-level image features extraction with textual description of image semantics.

The disadvantages of IRONS system are the presence of errors during curve evolution process and the restrictions for input visual queries, which must have small number of well-defined and separated objects. Additionally, significant occlusions between objects, week borders or complex background in image are not recommended in this application. The analysis of factors like tolerance to occlusion and deformation, robustness against noise, and feasibility of indexing are considered as possible extension of the proposed approach.

Acknowledgment

This research is sponsored by Mexican National Council of Science and Technology, CONACyT, Project #52333.

References

[1] O. Starostenko, A. Chávez-Aragón, G. Burlak, R. Contreras, “A Novel Star Field Approach for Shape Indexing in CBIR System”, *J. of Eng. Letters*, vol. , Oct. 2007, pp. 10-21.

[2] Q. Iqbal, Content Based Image REtrieval System, Univ. of Texas at Austin, Ph.D, 2007, http://amazon-ece.utexas.edu/~qasim/research.htm

[3] A. Chávez-Aragon, O. Starostenko, L. Flores-Pulido, “Star Fields: Improvements in Shape-Based Image Retrieval”, *J. Research on Comp. Science*, vol. 27, 2007, p.79-90

[4] T. Gevers, A. W. Smeulders, “Combining color and shape invariant features for image retrieval”, *IEEE Trans. on Image Proces*., vol. 9, (1), 2000, pp. 102-119.

[5] J.R. Ohm, F.B. Bunjamin, W. Liebsch, B. Makai, K. Muller, A set of visual feature descriptors and their combination in a low-level description scheme, J. Signal Process. Image Commun., vol. 16, 2000, pp. 157–179.

[6] L. Jia, J.Z. Wang. SIMPLIcity. (Semantics-sensitive Integrates Mathching). Stanford University 2000, http://wang14.ist.psu.edu/cgi-bin/zwang/regionsearch_show.cgi

[7] L.Flores, L. Altamirano, Classification of segmented images combining neural networks and wavelet matching, Machine vision, SPIE’s Photonics West, 2001.

[8] A. Del Bimbo, P. Pala, “Visual image retrieval by elastic matching of user sketche”, *IEEE Trans. Pattern Analysis Mach. Intell.* vol. 19 (2), 1997, pp. 121–132.

[9] F. Mokhtarian, “A theory of multiscale, curvature-based shape representation for planar curves”, *IEEE Trans. On Pattern Analysis Mach Intell.*, vol. 14 (8), 1992, pp. 789-805.

[10] M. S. Lew, *Principles of visual information retrieval, Advances in pattern recognition*, USA: Springer-Verlag, 2001.

[11] E. Jacobs, A. Finkelshtain, Fast Multiresolution Image Querying. SIGGRAPH, Conf. N.Y.1995.

[12] L. Flores, A. Lopez, L. Altamirano, Segmented Image Classification Combining Neural Networks and Wavelets, XXII Int. EEE Conf., 2000, pp. 23-27.

[13] O. Starostenko, A. Chávez-Aragón, “A Hybrid Approach for Image Retrieval with Ontological Content-Based Indexing”, *Lecture Notes, Progress in Pattern recognition*, Springer-Verlag, vol.. 3773, 2005, pp. 997-1004.

[14] D. Fensel, *Ontologies: a silver bullet for knowledge management and electronic commerce*, USA: Springer, 2nd ed., 2004.

[15] D. Beckett, “The design and implementation of the Redland RDF application framework”, *10th Int. WWW Conf*., 2001, pp. 120-125.

[16] S. Nayar, S. None, Columbia Object Image Library (COIL_100). Technical report, Columbia University. 1996.

Formal Specification and Simulation of the Robot Perceptual System

M. Yassine Belkhouche and Boumediene Belkhouche
Computer Science and Engineering Department
University of North Texas, Denton, Texas
myb0012@unt.edu
College of Information Technology, UAE University, Al-Ain, UAE
b.belkhouche@uaeu.ac.ae

***Abstract*- We use hybrid specification methods to capture the behavior of the perceptual system of autonomous mobile robots. A robot is divided into three main subsystems: the perceptual subsystem, the planning subsystem, and the acting subsystem. The perceptual system is itself decomposed into subsystems. We elaborate a formal description of these subsystems, and we model the communication protocol among them. A software simulation is developed to illustrate the behavior of the proposed design. The resulting model allows developers to test, verify, and simulate autonomous robot systems.**

***Index terms*—Formal specification, hybrid process algebras, hybrid automata, communication processes.**

I. Introduction

Formal specification of the behavior of hybrid systems, receives a great research interest in the recent years. Formal models such as hybrid automata, and hybrid process algebra has been developed for such purpose. Hybrid automata describe the system behavior using a directed graph, where the vertices represent the control modes, and edges represent switches between different control modes. Hybrid process algebra is another mathematical theory developed for the specification of hybrid system behavior. Hybrid process algebra uses algebraic expressions to describe the behavior of the system. A mobile robot is hardware/software system in which the control software plays a central role. Formally specifying such a system is a complex and critical task that requires the integration of continuous and discrete modeling. To decouple the complexity of the control tasks, the software model for the robot is decomposed into functional layers, each supervising a basic activity [1]. Higher level layers subsume the role of lower levels by suppressing their outputs. The functionalities of lower levels are not affected by the addition of higher layers. Layers dependence and communication are kept minimal to allow more extensibility.

In this research, we develop a formal model and a detailed specification of the behavior of an autonomous mobile robot designed to operate in complex environments [2]. The robot system is divided into three main subsystems: the perceptual subsystem, the planning subsystem, and the acting subsystem. An independent formal specification is elaborated for each subsystem. A composition and a communication protocol among these subsystems are used to formulate the complete model. Software simulation is developed to demonstrate the effectiveness and the validity of the proposed model.

For the purpose of this paper, we limit our description of the robot system model to just an overview. However, to illustrate our approach, we provide a detailed formal description of the perceptual system.

II. Description of a Wheeled Mobile Robot System

The software design shown in figure 1 illustrates the structure and the major components of the autonomous robot model. Figure 2 shows the architecture of the robot perceptual system. A robot consists of the following components:

- Coordinator: used to coordinate the activities among various subsystems.
- Perceptual system: used to sense the environment and to provide the robot with environmental context necessary for navigation.
- Path planning system: used to compute the navigation paths to move the robot from an initial position to its goal.
- Acting system: used to effect the motion and the steering of the robot.

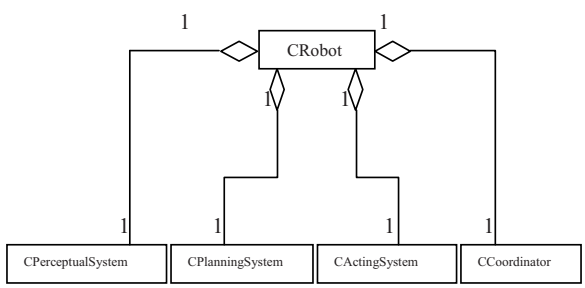


Figure1. Robot structure.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 140–143.

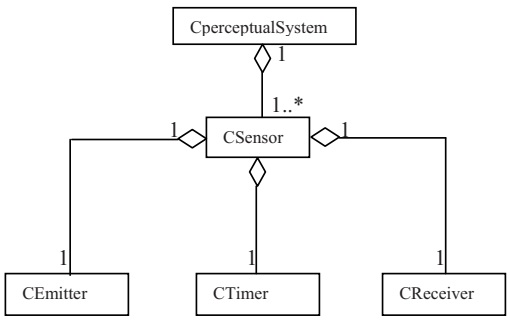


Figure2. Perceptual system architecture

III. FORMAL SPECIFICATION OF THE SONAR

A) *Process Algebra*

In this section a brief description of the basics, and most important operators needed to describe the behavior of the perceptual system of the robot is given [3, 4, 5]. Let P_1 and P_2 be two processes:

- The undelayable action *a*, is the process that performs the action *a* at the current point of time and then terminate successfully.
- We define the sequential composition of two processes P_1 and P_2 written $P_1 . P_2$, is the process that first behaves like P_1, but when P_1 terminates successfully it continues by behaving like P_2. That is, P_1 is followed by P_2. If P_1 never terminate successfully, the sequential composition of P_1 and P_2 will behave like P_1.
- The alternative composition of P_1 and P_2 denoted as P_1+P_2 is the process that represent the choice between the two processes P_1 and P_2. In other words, there is an arbitrary choice between P_1 and P_2. The choice is resolved on one of them performing its first action. The choice between idling processes will always be postponed until at least one of the processes can perform its first action.
- Another important process is the recursive process, used to describe the infinite behavior of a system. A recursive equation is said to be guarded if it is of the form $X=a.X$ where *a* is an atomic action.
- The most important operator for timing processes is the relative delay denoted by $a^t_{rel}(x)$, is the process that start executing *x* after the delay of *t* unit of time.
- The relative time out denoted by $w^t_{rel}(x)$, is the process that behaves either like the part of *x* that does not idle more than *t* units of time, or like the deadlocked process after a delay of *t* units of time if *x* is capable of idling for *t* units of time.
- The relative initialization of process *x* denoted by $\bar{v}^t_{rel}(x)$, is the process that behaves like the part of *x* that idles for *t* unit of time, if *x* is capable for this period of time, otherwise, it behaves like the deadlocked process after a delay of *t* unit of time.
- The parallel composition of two process written as $P_1 \| P_2$, is the process that chose to execute an initial action of P_1 or an initial action of P_2. The process $P_1 \| P_2$ can choose to execute a communication between transition of P_1 and P_2 to deal with this possibilities we assume a communication function $\gamma : Action \times Action \to Action$ which produces for each pair of *a* and *b* their communication $\gamma(a, b)$.
- The left merge written as $P_1 \lfloor\!\lfloor P_2$ takes it's initial action from the process P_1 and then behaves as the standard merge $\|$.
- The communication merge denoted by $P_1 | P_2$ executes as initial transition a communication between initial transition of the process P_1 and P_2, and then behaves as the standard merge $\|$.
- Another important notion is the trace of process. A trace of the behavior of process is a finite sequence of symbols recording the events in which the process has engaged up to some moment in time, we denote a trace by <*x, y, ...*> where *x* and *y* are atomic actions. Traces play a central role in recording, describing and understanding the behavior of processes.

B) *Sonar sensors*

Numerous and alternative sensor technologies have been developed for different kinds of robots. Sensors measure physical quantities like distance, light, sound, and temperature. Robot can be equipped with different types of sensors depending on the environment in which the robot is designed to operate in. Sonar sensors are widely used in robotic area for many reasons. Sonar sensors are devices that measure the distance between the sensor position and the obstacle. These kinds of sensors emit a powerful sonar pulse of a given range of frequencies, and it's reflected by the obstacle and detected by the sensor receiver. The time elapsed between the emission of the signal and the receiving of its reflected echo (time of flight) is used to compute the distance to the nearest obstacle to the sensor position. The sonar sensor can be viewed as the composition of three components: emitter, receiver and timer as shown in Figure 3. The communication between these three components defines the overall behavior of the robot sensing system. It is important to mention that sonar sensors are widely used in robotics to accomplish

different tasks such as navigation [6], localization, and mapping.

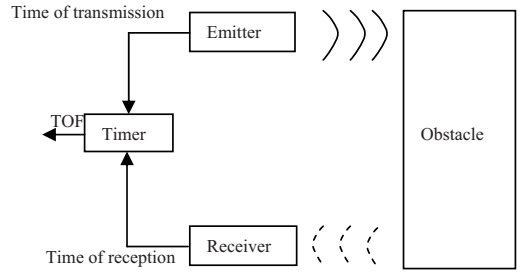


Figure3. Principle of the time-of-flight (TOF) sensors

In the following sections we describe the characteristics and the behavior of the robot perceptual system informally; later, the informal description is converted into a formal specification using mathematical equations.

C) Informal description

Before sending a sonar pulse the emitter sends a message to the timer to setup the timer at t_{max} (maximum value corresponding to the maximum distance that the sensor can detect an obstacle within it). If a reflected echo is received, the receiver sends a message to stop the timer. If no reflected echo is received the sensor returns the maximum value else the time of flight is calculated using the value returned by the timer. We suppose also that the time period at which the timer is set must be less than a given T_{timer} and the unit used in millisecond.

D) Formal description of the sonar sensors behavior

To describe the behavior of the sensor system describe by Figure 4, we define the following three processes, the process *SENSOR*, *EMITTER*, the process, *RECEIVER* and the process *TIMER* with the following set *A* of atomic actions:

A={emit(sig), receive(echo), setuptimer(t), send(exp), stoptimer(t), request(t)}.

emit(sig) : stand for emitting a sonar pulse.

receive(echo) : stand for receiving the reflected echo.

setuptimer(t) : stand for sending a message to the timer to setup this one at period of *t* units of time.

receive(echo) : stand for receiving the reflected echo.

send(exp): is the action of sending a notification of expiration.

stoptimer(t): stand for receiving a message from the receiver to stop the timer.

request(t): is the action of receiving a request from the emitter to set a timer at a time period of *t* units.

The mathematical specification of the robot sonar perceptual system is given by the following guarded recursive equations:

$$EMITTER = setuptimer(t_{max}).emit(sig).a_{\tau_{EE}}^{t_{max}}(EMITTER)$$

$$RECEIVER = receive(echo).stoptimer(t).RECEIVER$$

$$TIMER = \sum_{k<T_{timer}} request(k).\left(a_{\tau_{EE}}^{k}(TIMER) + a_{\tau_{EE}}^{k}(TIMER1)\right) + a_{\tau_{EE}}^{t}(TIMER)$$

$$TIMER1 = \sum_{k<T_{timer}} request(k).\left(a_{\tau_{EE}}^{k}(TIMER) + a_{\tau_{EE}}^{k}(TIMER1)\right) + a_{\tau_{EE}}^{t}(TIMER1) + send(exp)$$

$$SENSOR = EMITTER \parallel RECEIVER \parallel TIMER$$

E) Modeling communications

The communication between the processes *EMITTER*, *RECEIVER*, and *TIMER* is illustrated by the following graph:

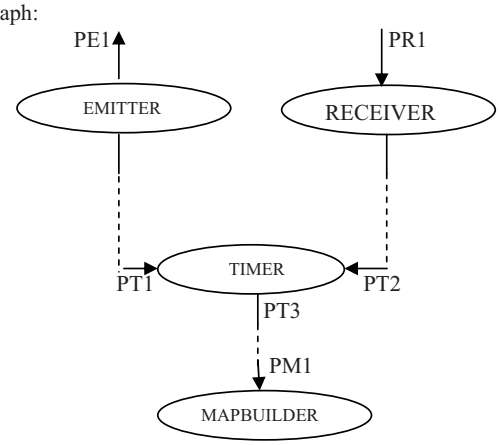


Figure4. Sonar communication model

IV. SIMULATION

The previous specifications were implemented as graphical simulation using C++ and OpenGL. Figure 5 shows the simulation of the sonar perceptual system.

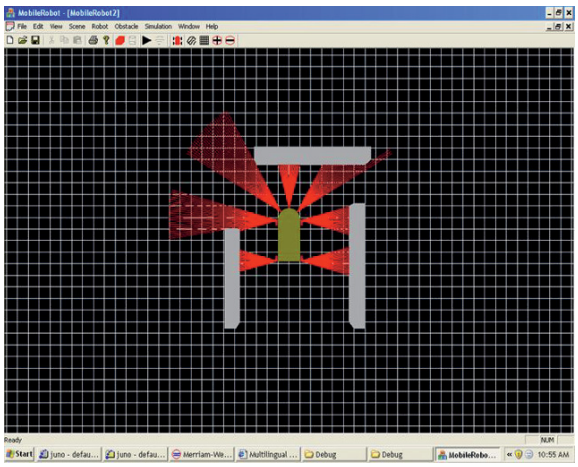


Figure5. Simulation of sonar sensors

V. CONCLUSION

Mathematical description of the behavior of hybrid systems plays a very important role in the design, implementation, testing, simulating and improving the system efficiency. In this paper we described the behavior of the robot sonar perceptual system, which allows us to combine the perceptual subsystem with the planning and acting modules to provide a complete behavior of the robot system. In our future research, we will consider other more complex perceptual systems such as vision for example, as well as other aspect in behavioral and humanoid robots.

REFERENCES

[1] Brooks, R. A, "A layered control system for a mobile robot", *IEEE Transactions on Robotic and Automation* **18** (2002), no.5, pp. 847-851.
[2] Belkhouche, F. and B. Belkhouche, "Wheeled mobile robot navigation using proportional navigation", Advanced Robotics (2007), pp 1-26
[3] Bergstra, J. and C. Middelburg, "Process algebra for hybrid systems", theoretical computer science 14 (2005), no.6, pp 215-280.
[4] Baeten, J. and C. Middelburg, "Process algebra with timing", Springer, Aug 26,2002.
[5] Cuipers, P. and M. Reniers, "Hybrid process algebra", Computer Sceince Report 03-07, Department of Mathematics and Computer Science Eindhoven University of Technology, July 2003.
[6] Leonard, J.J and H.F Durrant-Whyte, "Directed sonar sensing for mobile robot navigation", Kluwer, Dordrecht, the Netherlands, 1992

Enhancing Diagnosis Ability for Embedded Electronic Systems Using Co-Modeling

Manel KHLIF, Mohamed SHAWKY
Heudiasyc - UMR 6599
Université de Technologie de Compiègne
Centre de Recherches de Royallieu
BP 20529

***Abstract*- This paper describes a new modeling and simulation approach in order to enhance the diagnosis ability of an electronic embedded system, including in the automotive field. Our modeling approach integrates the hardware specifications to the functional model in order to establish better system observation. It is based on hardware/software (HW/SW) co-modeling with multilevel of granularity.**

To reach this objective, we have set up a relationship between the desired diagnosis accuracy and the level of granularity of the HW/SW co-model, for every Electronic Computing Unit (ECU). Our contribution allows the attribution of the right co-simulation hierarchical level by attributing the right simulation accuracy, for each function under observation.

***Index Terms*- hierarchical modeling, HW/SW co-modeling, real time simulation.**

I. Introduction

The technological development encourages the car manufacturers to propose advanced driving assistance functions that involve more than one computing unit. In fact, a computing unit uses information issued from sensors or other computing units, yielding a "system with distributed functions". In a vehicle, the functions are sometimes distributed on several components or subsystems (computing units, wires, sensors, actuators…), communicating with several interconnection networks [1].

However, one of the disadvantages of this distribution is the difficulty of the real time supervision to detect and localize a fault, especially electronic hardware faults. To bring out the advantages of a highly distributed architecture, we propose a modeling methodology that benefits from the existing link between the software and the hardware platforms.

Our contribution in this paper is to enrich the functional models with hardware characteristics, at the very first phase of establishing the diagnosis models (system observation), in order to reproduce the appropriate behavior of the system in a set of comprehensible models showing at the same time the hardware and the software behaviors.

To present our contributions, this paper is structured as follows:

First, we present the need of HW/SW co-modeling for embedded electronic systems supervision. Then, we present the related works done in the field of the HW/SW co-design [2].

In section IV, we show a relationship between the required accuracy for fault detection and the level of granularity in the HW/SW co-model, in order to find an appropriate compromise between fault detection accuracy and simulation speed. In section V, we use SystemC as a working environment for the hierarchical HW/SW co-modeling of our embedded electronic architecture. We present the results in sections VI and VII.

Finally, in the last section we conclude this paper and present our future works.

II. Hardware/Software Co-Modeling for Embedded Electronic Systems

Car manufacturers usually employ software models expressing the embedded functions to make fault detection and diagnosis. However, when the hardware architecture reaches a complex level of functional distribution, it becomes difficult to a diagnosis designer to maintain the HW/SW link for each function or sub-function of the system in the diagnosis model.

We believe that every sub-function has a link with at least one hardware sub-component, and a hardware fault appears in the system as a functional fault. Therefore, if we detect the functional fault we can localize the hardware fault if we know exactly the existing link between the sub-function and the hardware sub-component.

The electronic distributed architecture that we study is embedded on board of a truck. It is composed of a set of ECUs connected by the interconnection bus CAN (Controller Area Network) [3]. Every ECU is composed of a processor, a memory, a CAN interface and eventually Inputs/Outputs interfaces.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 144–149.

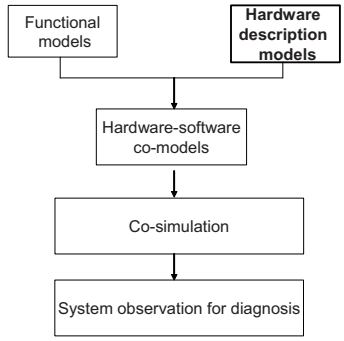


Fig. 1. Hardware and software models for system observation for diagnosis.

Some existing techniques of fault detection and diagnosis, especially of electronic systems, are based on functional simulation of the real controlled system, running on a prototyping platform, in parallel with the real time operation of the system [4]. These techniques do not describe hardware components (i.e. architecture and behavior description) and need to be more accurate by expressing simultaneously both hardware and software behaviors, in order to lead to more accurate results of fault detection (Fig. 1).

We aim at modeling the functional distribution on the hardware architecture, in a coherent way, offering the ability to be connected to other models that may be proposed by different manufacturers of such heterogeneous electronic system. Our approach should allow the supervision system designer to integrate different models to simulate and test the system.

In the next section we present research works close to our objectives, and that use mix hardware and software information for simulation.

III. HW/SW Co-design

Hardware/software co-design is a set of methodologies and techniques specifically created to support the concurrent design of both systems, effectively reducing multiple iterations and major redesigns [2].

Within the context of co-design methodologies, concurrent hardware and software techniques have been proposed in the literature employing for example SpecC to add more details at the specification level [5]. The adoption of various formal languages for co-simulation, like SDL and C [6] is mainly used for the design of reactive systems like telecom systems such as wireless protocols employing different and standardized formalisms. On the other hand, functional models compatible with HDL (Hardware Description Language) models [7] are needed to get accurate hardware specifications using for example RTL (Register Transfer Level) level of modeling. However when it concerns HW/SW co-design for model based diagnosis, system's behavior and properties are specified in a single formal language such as in [8] and [9].

Even though Hardware/Software co-design is receiving a lot of attention in literature, most published works do not address model-based diagnosis for electronic systems. In this paper, we are not interested by dependability analysis (fault simulation, estimation of optimal diagnostic strategies, etc,), but we focus on co-modeling of the appropriate behavior of an electronic system for co-simulation oriented diagnosis.

IV. Multilevel of granularity co-modeling

Multilevel of granularity is a hierarchical view of a system, expressing in each level a degree of details and accuracy.

Our objective is also to develop a relationship between the accuracy of the expected on-line fault detection and the level of granularity of hardware/software co-modeling (Fig. 2).

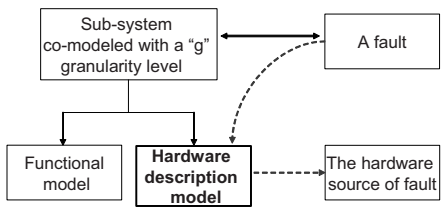


Fig. 2. Relationship between a sub-system HW/SW co-model and hardware source of fault.

As a first step, we have to model the hardware architecture as a set of hardware sub-components. Then, we have to model the software platform as a set of sub-functions allocating them to the modeled hardware sub-systems.

As a second step, we define a scale of criticality levels for the sub-functions. Then we allocate a granularity level of modeling to every criticality level, and hence, each sub-function is co-modeled with a corresponding granularity level (Fig. 3).

The more accurate the level of granularity, the longer the simulation time is. Thus, it is possible to switch between two or more levels of granularity according to the criticality level of eventual faults and to the diagnosis system needs (e.g.: functions priorities). This is the main advantage of our approach of modeling at various levels of granularity.

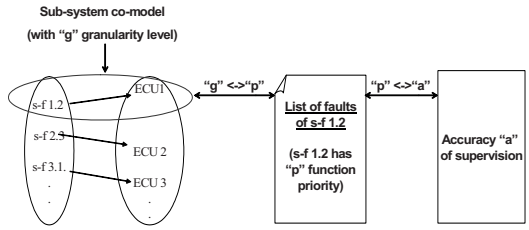


Fig. 3. Granularity-accuracy relationship.

V. WORKING ENVIRONMENT

A. SystemC

One of the most promising SystemC advantages is HW/SW co-modeling to develop virtual platforms, because it supports a unified language of HW/SW modeling [10].

We have selected SystemC as a working environment because it has many advantages:

- SystemC allows hierarchical modeling to express the multilevel of granularity modeling
- It allows HW/SW modeling with the same language
- The models could be easily connected to any other hardware models [11], or functional models (e.g. in Simulink) [12][13][14]
- SystemC environment includes also a simulator: it consists of a C++ library and an event-based motor for simulation
- Any C or C++ library can be included in a HW/SW co-model

Hence, we can describe the appropriate behavior of the electronic embedded system with different levels of hierarchy. Thus, every sub-system that should be under supervision can be hierarchically co-modeled.

B. Cycle Accurate modeling

SystemC Transaction-level modeling (TLM) is a high-level approach to model digital systems where details of communication among modules are separated from the details of the implementation of functional units or of the communication architecture [15].

A Cycle Accurate (CA) model is a TLM model that represents the stage of communication refinement, in which communication is modeled accurately down to the level of bus cycles, and even clock cycles. CA modeling allows hardware verification, evaluating the state of every component in every cycle and running software device drivers. CA simulation speed varies between 10 and 100 KHz. A CA model consists of a set of processes that run once per cycle. This fits with the use of SC_METHOD processes and non-blocking calls.

Fig. 4 shows that TLM projects do not require a lot of effort and time to be correctly modeled compared to RTL projects. In fact, a Cycle Accurate project may need approximately half of the time compared to an RTL project for its realization.

For these advantages, we have used the CA level to co-model our HW/SW platform as shown in the next section.

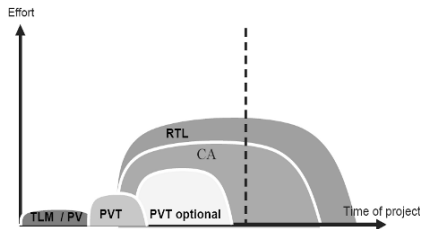


Fig. 4. Time and effort spent for RTL and TLM use cases [16].

VI. MULTILEVEL CO-MODELS

A. Multilevel of granularity using TLM modeling

We co-modeled with TLM, hierarchically, each HW/SW sub-system beginning with the highest level of granularity. Thus, on each level of granularity, we find a set of models representing at the same time the functional behavior and the hardware architecture.

In the next sub-sections we show an example of TLM model representing the embedded electronic architecture.

B. ECUs and CAN bus modeling

The whole architecture consists of n ECUs communicating through the CAN network [3]. In this part of the work, we have modeled the CAN protocol real-time behavior to realize communications between ECUs models. We have simplified the details to ease the modeling; by implementing a virtual arbiter in the bus. With the Transaction modeling, the communication between components is described as function calls.

Each ECU is master and slave at the same time and has one bidirectional port in each module. It is used to send orders to the bus (Requests) and getting data and information from the bus (Responses) (see Fig. 5). Each ECU that wants to send a message sends a request to the bus. If at least 2 ECUs request a bus transmission at the same time (i.e. in a time shorter than a bus cycle), the bus arbiter selects the most important message by comparing arbitration fields in the two messages.

Only one clock is used for all processors when the level of granularity is high and the accuracy of the model for simulation is set to the ECU clock cycle.

It is important to note that full CAN protocol is used only in models with high level of granularity, expressing transactions between ECUs. With a more accurate level of granularity, the

processor and the memory models of every ECU are wrapped into SystemC modules in order to communicate with other devices such as CAN controllers. Anyway, the transactions are kept cycle accurate.

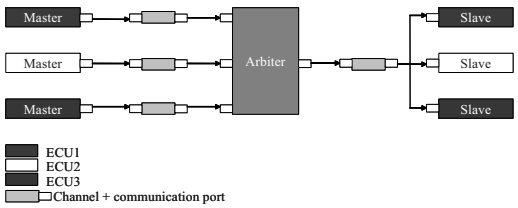


Fig. 5. CAN bus TLM model.

C. HW/SW models interfacing

The communication between hardware modules and software modules should be done through message exchange. Messages should be exchanged through a shared interface between the two sides, implementing the needed methods for message exchange and eventually using blocking calls (Fig. 6).

We have co-modeled in SystemC a sub-function of Lane Keeping [17]; a given function from a truck manufacturer. The Lane Keeping should keep a truck centered in the current lane of a highway, detected by an embedded camera. Fig. 6 shows also the example of the embedded camera software that sends the acquired videos to an ECU (the hardware module), which is responsible of filtering all extra images from the video sequence, to get only the video sequence of the highway.

D. Interfacing SystemC co-models

A systemC co-model may be the result of a higher level system level modeling, such as SysML [18] or AADL (Architecture Analysis and Design Language) [19]. A SystemC co-model can also integrate other functional models in Simulink [12] for SW description and VHDL [20] for HW description models. We show also that co-simulation is possible by using interfaces between the heterogeneous simulators (Fig. 7). For example, [11] presents some results of design practice of HW modules; co-simulation and synthesis are combined to achieve higher abstraction levels in the design. The SystemC-VHDL co-simulator tool is also based on a SystemC/C++ front-end developed to support the co-simulation between VHDL and SystemC. Another example of co-simulation tool is shown in [13] using Simulink with SystemC in a cycle-accurate context.

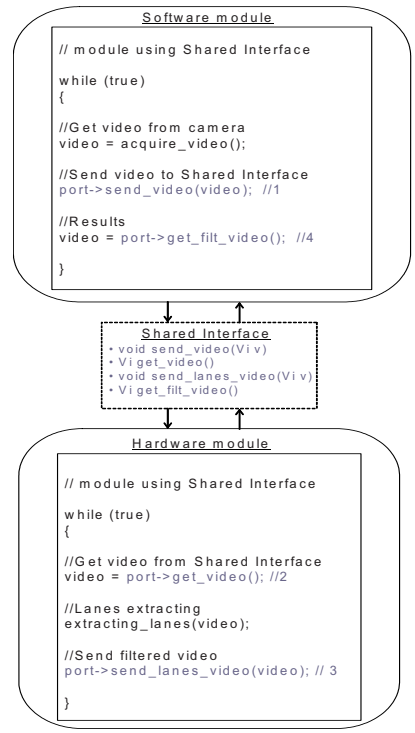


Fig. 6. HW/SW modules interfacing.

The advantage of the proposed multilevel granularity co-modeling is the possibility of establishing a link between a HW/SW co-model and another type of models in any level of detail and having several possibilities of co-simulation speed for fault detection.

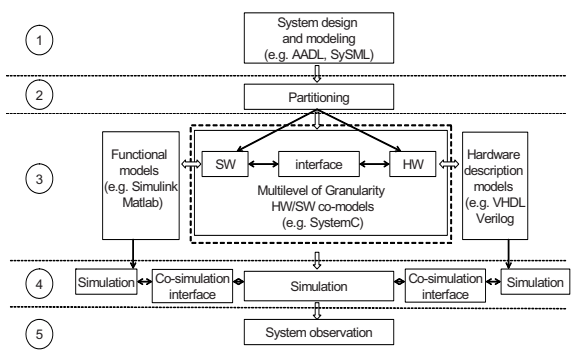


Fig. 7. SystemC co-models communication with different simulation tools.

VII. CO-SIMULATION FOR DIAGNOSIS

The supervisory system can simulate the co-models beginning with the highest level of the hierarchy in order to observe the electronic system. In case of incoherence detection in a given level of the hierarchy, our approach brings the advantage of co-simulating the system with a more accurate level in order to increase the accuracy of incoherence localization (Fig. 8).

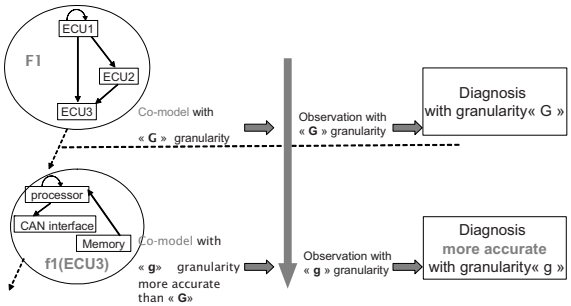


Fig. 8. Multilevel of granularity co-simulation.

It is important to add that this approach of modeling for diagnosis is simple to implement when the embedded functions are cyclic, as progressive and repetitive simulations may be done with real input values. For this reason, we have tested our approach with the Lane Keeping (LK) system described earlier (Fig. 9).

This system supplies a wheel angle set point sent to the ECUs, and does not require any driver intervention. At any time, the driver can deactivate the LK function. The LK system recognizes the lane markings using an embedded camera.

In case of fault detection with a high level of granularity co-simulation, the diagnosis system should make sure that the electronic embedded system may suffers from a fault by co-simulating and testing the same sub-function (e.g. Highway videos capturing) with a more accurate level of granularity until reaching the sub-system eventual source of fault.

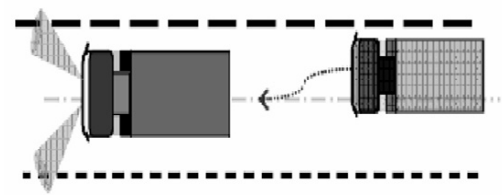


Fig. 9. Lane Keeping function [14]

VIII. CONCLUSIONS AND PERSPECTIVES

This paper has presented a co-modeling technique with multilevel of granularity under SystemC for electronic systems. Cycle accurate models of the CAN bus, CAN interfaces and ECUs including memories and processors has been co-modeled with different levels of granularity.

The impact of a multi-granularity HW/SW model for the hardware real-time fault detection has been shown.

This result shows that such modeling method is suitable to simulate distributed real time HW/SW architecture, especially electronic architectures and allows the supervision designer to re-use the co-models in embedded simulations for real time supervision or to simulate and test the system off-line.

As future work, we aim to extend the multilevel granularity modeling to other computing platforms, like SoC (System On Chip) for example; a technology which is gaining interest for car manufacturers. The TLM is very suitable to model the architecture of a SoC, in order to enable development of the embedded software in advance of the hardware, and to carry out analyses earlier in the design cycle.

ACKNOWLEDGMENT

We would like to thank Amanie Ghannoum and Janine Alhassan for their contribution to this work.

REFERENCES

[1] Paret D. 2007. Multiplexed Networks for Embedded Systems: CAN, LIN, FlexRay, Safe-by-Wire, *Wiley*, ISBN: 978-0-470-03416-3
[2] Giovanni. De Micheli, Rolf. Ernst. Readings in hardware/software co-design. 2001. *Embedded System Computer*
[3] Paret D, 1999. Le bus CAN description : de la théorie à la pratique. *Dunod* , ISBN-10: 2100047647, ISBN-13: 978-2100047642.
[4] Hamscher W. and al, 1992. Readings in model-based diagnosis. *Morgan Kaufmann,* isbn: 1-55860-249-6, San Francisco, CA, USA.
[5] Gajski D-D. and al, 2000. SpecC: Specification Language and Methodlogy, *Kluwer Academic Publishers*.
[6] Gioulekas, F. and al, 2005. Heterogeneous system level co-simulation for the design of telecommunication systems. *Journal of Systems Architecture. 51*, p. 688-705.
[7] Wong, S.Y, 1998. Hardware/software co-design language compatible with VHDL. *IEEE, WESCON/98*, Anaheim, CA, USA, p. 78 -83.
[8] Csertan G. and al, 1994. Modeling of Fault-Tolerant Computing Systems. *In Proceedings of the 8th Symposium on Microcomputers and Applications, uP'94*, Budapest, Hungary, p.78-83.
[9] Csertan G. and al, 1995. Dependability Analysis in HW/SW co-design. In Proceedings of the IEEE International Computer Performance and Dependability Symposium, IPDS'95, Erlangen, Germany, p. 316-325.
[10] Grötker T. and al, 2002. System Design with SystemC. Springer, Chapter 8, p. 131. ISBN 1402070721.
[11] Bombana, M.; Bruschi, F. 2003. SystemC-VHDL co-simulation and synthesis in the HW domain. *Design, Automation and Test in Europe Conference and Exhibition*, pp. 101-105, Messe Munich, Germany.
[12] Warwick C, "SystemC calls MATLAB", MATLAB Central, March 2003, http://www.mathworks.com/matlabcentral/
[13] Czerner F. And Zellmann J. 2002. Modeling Cycle-Accurate Hardware with Matlab/Simulink using SystemC. *6th European SystemC Users Group Meeting (ESCUG)*. Stresa, Ilalia.
[14] Boland J-F and al. 2004. Using Matlab and Simulink in a SystemC verification Environment. *2nd North American SystemC User's Group*. Santa Clara, CA, USA

[15] Cai L. and Daniel G, 2003. Transaction level modeling: An overview. In Hardware/Software Co design and System Synthesis, *Report 03-10, Center for Embedded Computer Systems,* University of California, p. 466-471.

[16] Ghenassia, F. (ed.), 2005. Transaction-Level Modeling with SystemC. TLM Concepts and Applications for Embedded Systems. *Springer*. ISBN 0-387-26232-6.

[17] Claeys X. and al, 2003. Chauffeur Assistant Functions. *Report restricted to RENAULT TRUCKS,* Contract number IST-1999-10048, Lyon, FRANCE.

[18] SysML Partners. 2005. Systems Modeling Language (SysML) Specification. Version 1.0 alpha. *SysML.org*

[19] http://www.axlog.fr/aadl/http://www.mathworks.com/products/simulink/

[20] Cote, C.; Zilic, Z. 2002. Automated SystemC to VHDL translation in hardware/software co-design. *Electronics, Circuits and Systems, 2002. 9th International Conference on Volume 2, Issue*, pp. 717-720, Dubrovnik Croatia.

Development Environment Using FPGA for Domotics Applications Based on X10 Technology

Manuel D. Cruz[1], Juan A. Ortega[1], Ángel Barriga[2] and Alejandro Fernández-Montes[1]

[1] Department of Languages and Computer Systems, University of Seville, Spain.
[2] IMSE CNM-CSIC / University of Seville, Spain.

***Abstract*—This communication proposes a basic software and hardware architecture of a controller for the X10 technology interface CM11A, oriented to the world of home automation. The implementation of the system is based in the use of programmable devices such as FPGA (Field Programmable Gate Array). With this controller an end user will be able to control and to manage a set of devices distributed in a domotics space.**

I. Introduction

In order to find home automation low-cost and little-invasive technological solutions, and easy handling for the end user, the use of FPGA embedded systems will be an implementation option to have into account [1]. Home automation or domotics are automation technologies applied to buildings. This automation technologies use control devices, sensors and actuators. These devices communicate between them by different communication schemes such as specific buses, wireless communication, or electrical network. From end of the nineties there appeared different standards of communication like CEBus, LonWorks, KNX [2, 3].

The use of the electrical network facilitates the installation of these technologies in the buildings since it is not needed of any additional communication infrastructure. In this sense, devices and communications protocols adapted to this media have been developed [4, 5]. One of the most widespread is the X10 protocol [5, 6]. This one was developed in 1978 by engineers of the company Pico Electronics Ltd. in Glenrothes, Scotland. Since then the technologies based on the utilization of the electrical network known by Power Line Carrier (PLC) have been spreading, appearing a great variety of commercial products and standards [7].

In the area of home automation architectures focused on the education, the systems based on FPGA open a broad field of possibilities [8]. The present communication focuses the process of implementation of a controller capable of communicating with all the devices connected to an X10 network distributed in a domotics space. In particular the developed system focuses in an application to the illumination control. With the control of the illumination integrated in a domotics system it is possible to obtain an important power saving and an increase of the comfort. The next section describes briefly the X10 protocol. Later the FPGA platform is presented. Finally, the last section describes the design of the system.

II. The X10 Protocol

The X10 communication protocol allows controlling remote devices using the electrical network and the specific modules to which they are connected. The control signals are based on the transmission of RF pulses (120 KHz). The transmissions are synchronized with the crossing by zero of the alternating current (AC). The binary one is represented by a pulse of 120 KHz. during 1 millisecond whereas binary zero is represented by the lack of this 120 KHz pulse.

The protocol [9] consists of address bits and commands. In an X10 network can coexist up to 256 different interconnected modules. The way of identifying them is by a code (*housecode*) composed by a letter (from A to P) and by a number (from 1 to 16, the device code). These modules can recognize (depending on its characteristics.) up to 16 different operation codes.

An X10 transmission from the controller, implemented in a development board (DB), to the X10 interface consists of the communication of a *housecode* and device code combination followed by the command code. The sequence of the transmission can be seen in the figure 1. We can observe what each one of the frames of bytes means. This format of transmission is always the same regardless of command or address.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 150–153.

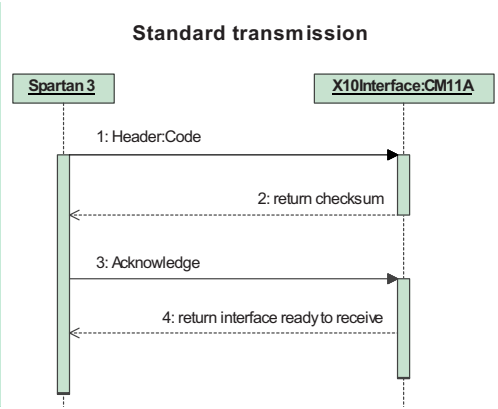


Fig. 1. X10 standard transmission

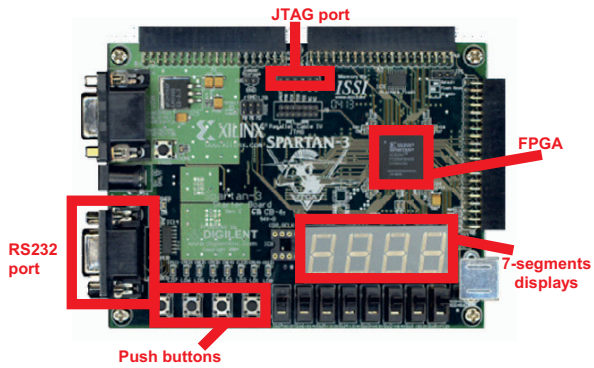


Fig. 2. Xilinx's Spartan-3 development board

1. The DB transmit the *header:code/command* to the X10 interface.
2. The X10 interface receives the frame and makes a checksum. It transmits it back, to the DB.
3. The DB checks the checksum and sends an acknowledgement (ACK) frame.
4. The interface receives the ACK and informs that it is ready for more transmissions.

We can observe a code example in the System Design section, subsection B.

There are an extensive range of products based on X10 protocol which can be classified in the following categories: controllers, modules and complements. The system describe in this paper is composed by a CM11A model PC programmer (belonging to the category of the controllers.) and a lamp module (that receives commands from the electrical network and acts on a lamp.).

III. Hardware Development Platform

The control system has been implemented on an FPGA development board containing a Xilinx Spartan-3 [10]. The system operates with a clock frequency of 50 MHz determined by an oscillator included on the board. The board incorporates a set of elements that facilitate the system development and allows diverse applications. It has two SRAM memory modules up to 256 KB that store the application. The RS232 port allows the communication with the X10 interface.

The address (*housecode*) and the command (*operationcode*) are formed activating four push-buttons. The 7-segments displays allow verifying the operation and inform, as well, about the state of the controller or the X10 devices. Finally, the JTAG communication port has been used for download and test the application in the FPGA In figure 2 the components that have been used are shown.

The FPGA development environment is the Xilinx Platform Studio (XPS), of which we have used the tools that allow covering the design flow of embedded systems. This flow consists of the hardware architecture description stage, the synthesis and implementation of the control circuit, the development of the software applications, the compilation of the above mentioned applications and the programming of the FPGA and the SRAM memory.

The FPGA board controls any X10 receiver module in the home by means of the computer interface module CM11A (figure 3a). The CM11A plugs into an outlet and connects to the RS232 serial port on the FPGA board

The lamp module, shown in figure 3b, is used to control a lamp through X10 protocol. It is controlled by the remote CM11A module. It connects the existing lamp in the socket of this module, which is connected in the socket of the wall. It is the interface module between the lamp and the socket of the wall without requiring any additional connection and any modification in the existing electrical circuit. It offers the possibility of switch on/off the lamp and varying its luminosity remotely by means of an X10 controller.

IV. System Design

A. Hardware architecture

Figure 4 illustrate the configuration scheme of the system. In this scheme the controller has been implemented on the FPGA development board described in the previous section. This board communicates with the X10 interface module by means of the RS232 port. This one transmits the address and commands through the electrical network. The X10 lamp module that has been address receives the commands and acts on the device that it has connected.

a) b)

Fig. 3. a) CM11A X10 computer interface module, b) lamp module

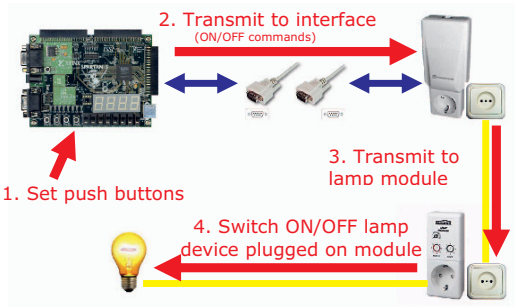


Fig. 4. Communications scheme

The design of the controller has been realized applying a hardware&software codesign methodology that has allowed defining an open and flexible architecture. The architecture is open in the sense that allows incorporating new elements and it is flexible because it allows to be reconfigured.

The system is constituted by two types of elements: the hardware components that constitute the circuit programmed in the FPGA, and the application software that runs on the hardware and is stored in the system memory. The partition of the functionality of the system into the hardware and software components has been realized having in account the already mentioned of open and flexible architecture. For it the circuit has been developed with IP (Intellectual Property) modules. These IP modules allow the reuse and easy the insertion of new functionality.

The architecture, which shows in figure 5, is based on the utilization of the MicroBlaze processor from Xilinx. The MicroBlaze processor is a soft-core 32 bits RISC architecture that can be include in Xilinx's FPGA as an IP module. MicroBlaze has Harvard architecture; this means separated data and addresses buses. It has several structures of buses of which we have used the LMB (Local Memory Bus) and the OPB (Peripheral On-chip Bus). Bus LMB allows accessing the internal FPGA memory in only a clock cycle. The disadvantage is in that the size of this memory (BRAM) is limited to the available resources in the FPGA device.

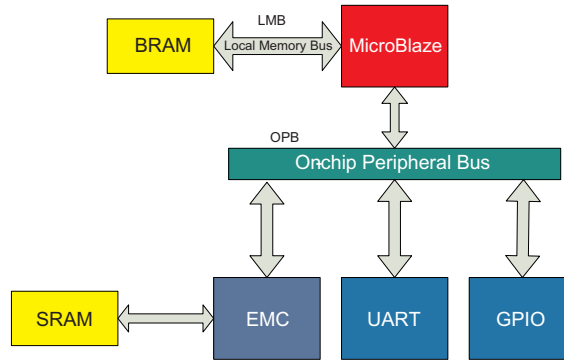


Fig. 5. Controller architecture

The processor communicates with the peripheral devices by an OPB (On-chip Peripheral Bus) IBM's standard bus. The peripheral ones are mapped in the memory address space. This means that the inclusion of new peripheral consists simply of connecting it to the OPB bus and including it within the memory address space of the processor. In our application only the three devices shown in figure 5 have been required. The peripheral devices correspond to a parallel port (*gpio*), a serial port (*uart*) and a memory controller (*emc*). The parallel port receives information of the push-buttons of the board and generates the signals to the 7-segment displays. The serial port allows realizing the transmission of information towards the X10 module. The external memory controller facilitates the transfer of information with the SRAM memory that contains the application software.

In figure 5 it is possible to observe that the system has two RAM memory blocks: BRAM and SRAM. BRAM is the FPGA internal RAM memory whereas the SRAM is the FPGA external memory module. The employment of both memories justifies itself due to the fact of the size of the software application requirements. BRAM module has better access time (one clock cycle), nevertheless its size is limited. Block SRAM can be expanded based on the system requirements. There are three expansion connectors in the development board to adapt any additional device such as a memory module.

B. Software application development

The software application has been implemented using C++ language. The XPS development environment provides a series of software packages developed in ANSI C that facilitate the codification of the user's applications (peripheral drivers, libraries and operating systems) [11].

Figure 6 shows the model-view-controller (MVC) of the software application describing the classes' graph. The pattern has been used in order to separate data and user interface. In that way it is possible to improve and expand the application in an easy manner. This model divides the application in three layers:

- View: it is the layer that transmits and receives information of the end user. In order to include new services it is necessary to add more classes that implement the interconnection between the final user and the devices. Two classes belong to this layer:
  DisplayDriverClass: it is the class that shows the results and information of interest in the 7-segment display.
  ButtonsDriverClass: it is the class that receives the push-buttons signals and transmits the information to the controller.

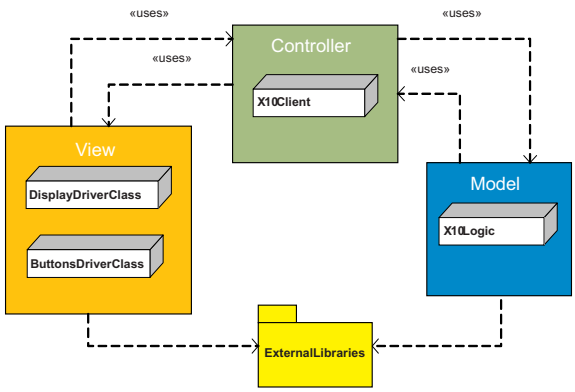


Fig. 6. Model-view-controller.

- Controller: in this layer the information is processed and transmitted up to the other two layers. For example, it receives the information of the pressed button and transmits the orders to the model layer, which will transmit the command to the target device through the serial port.
  X10Client: this class works as an adapter, reducing responsibility to the classes of the other two layers and improving the model of the system. This way a major independence is obtained between all the classes facilitating later modifications.
- Model: in this layer the whole application logic is carried out. It is the one that establish the communication with the X10 devices through the serial port.
  X10Logic: it is the class that contains all the set of necessary methods to establish the communication with the X10 interface through the serial port, among other functionality, there is the communication protocol of this technology. Figure 7 shows the code of the typical interaction between X10 interface and the development board.

CONCLUSIONS

This communication describes a system capable of controlling the devices of X10 technology distributed in a domotics space using an FPGA device. One of the objectives has been that the system is expandable and modifiable, so that it is possible in the future to add new functionalities and services to the end user. This will allow to incorporate new external devices in order to increase the functionality and improving the user interface (numeric keypads, LCD screens, tactile screens, etc). The system presented in this communication constitutes a development platform focused on economic solutions and easy managing in the world of the home automation

```
…
//next transmission, it send the address A1
outputBuffer = _HEADER_;//header
while (!XUartLite_Send  (&uart, &outputBuffer, 1));

outputBuffer = _DEVICE_ADDRESS;//device address
while (!XUartLite_Send  (&uart, &outputBuffer, 1));

while (!XUartLite_Recv  (&uart, &inputBuffer, 1));
//check the checksum
if (inputBuffer == ((_HEADER_ + _DEVICE_ADDRESS)&0xff))
{
   //checksum ok
   //transmit ACK to the board
   outputBuffer = _OK_;
   while (!XUartLite_Send (&uart, &outputBuffer, 1));

   while (!XUartLite_Recv  (&uart, &inputBuffer, 1));
   for (j=0; j<20000000; j++);//wait for synchronize

   if (inputBuffer == 0x55)
   {
      //interface ready for the next transmission
      …
```

Fig. 7. Code example of the X10 transmission

ACKNOWLEDGEMENT

This work was supported in part by the Spanish Education and Research Council under grants no. TEC2005-04359/MIC and no. TSI2006-13390-C02-02, and by the Andalusia Regional Government under grants no. TIC2006-635 and no. TIC2141.

REFERENCES

[1] Renato Nunes, "Implementing Tiny Embedded Systems with Networking Capabilities", IADIS International Conference on Applied Computing 2005, Algarve, Portugal, February 2005.
[2] LonWorks (ANSI/EIA 709.1-A), http://www.echelon.com.
[3] KNX, http://www.knx.org/.
[4] HomePlug, http://www.homeplug.org/home.
[5] R. N. Bucceri, "The Latest Technology in Automated Home Control - Book System Design Manual Using X-10 & Hardwired Protocols", Silent Servant, Inc, 2003.
[6] X10, http://www.x10.com
[7] Universal Powerline Bus, http://www.pcslighting.com/upb/overview.html.
[8] F. Mateos, V. M. González , R. Poo, M. García, R. Olaiz, "Design and Development of an Automatic Small-Scale House for Teaching Domotics", 31st ASEE/IEEE Frontiers in Education Conference, Reno, NV-USA, 2001.
[9] X10, CM11A Interface Communication Protocol, ftp://ftp.x10.com/pub/manuals/cm11a_protocol.txt.
[10] Xilinx, Inc. Spartan-3 Starter Kit Board User Guide, http://www.xilinx.com/bvdocs/userguides/ug130.pdf.
[11] P. Anderson. "Xilinx Platform Studio Tutorial". Embedded Systems Design- Advanced Course Homepage. http://www.cs.lth.se/EDA385/. 2006.

Robustness of a Robot Control Scheme for Liquid Transfer

M. P. Tzamtzi
Department of Automation
Halkis Institute of Technology
34400 Psahna, Evia, Greece
tzamtzi@teihal.gr.

F. N. Koumboulis
Department of Automation
Halkis Institute of Technology
34400 Psahna, Evia, Greece
koumboulis@teihal.gr.

Abstract- **The robust performance of a control design scheme for sloshing suppression and container positioning during liquid transfer using robotic manipulators is studied with respect to parametric uncertainty. The control scheme combines a partial inverse dynamics controller with a heuristically tuned PID controller. Robustness is studied with respect to uncertainties on the parameters of the liquid, namely the liquid's density and viscosity, as well as the liquid's level within the tank. The range of uncertainties that can be well tolerated without critical deterioration of the closed-loop performance is determined using simulation results. Moreover, an enhancement of the control design scheme is proposed that improves robustness.**

I. Introduction

Sloshing suppression during liquid transfer and pouring has been studied using several control approaches ([1]-[15]), as for example optimal control, H_∞ control methods, input shaping control, application of time varying filter gain and hybrid shape approach. Interest on sloshing suppression is raised due to its importance in several industrial applications, like casting and steel industries, where molten metal has usually to be moved from the furnace to the casting areas. Excessive sloshing may cause overflow, deterioration of the product quality due to contamination or excessive cooling of the molten metal [1].

In [15] a two stage control design scheme has been proposed for the case of a liquid container carried by an articulated manipulator (see Figure 1). The proposed scheme takes into account the dynamics of the manipulator. In order to simplify modeling of the liquid's motion within the tank, a well-known pendulum-type model is used ([1]-[5], [8], [10], [11], [13]-[15]). The proposed control scheme combines a partial inverse dynamics controller with a PID controller, which is tuned with the use of a "metaheuristic" search algorithm introduced in [16]. The design goal is to achieve desired positioning of the tank, while simultaneously reducing the amplitude and the duration of liquid's oscillations, which are indirectly controlled, through appropriate rotation of the tank actuated by the robot's end-effector.

The proposed controllers exploit measurements of the manipulator and the container position and velocity variables, but not measurements of the pendulum's position and velocity. However, the application of the metaheuristic search algorithm for tuning the PID controller requires knowledge of the pendulum model parameters, which are directly related to physical parameters of the liquid, namely the mass and the height of the liquid and an equivalent viscosity coefficient.

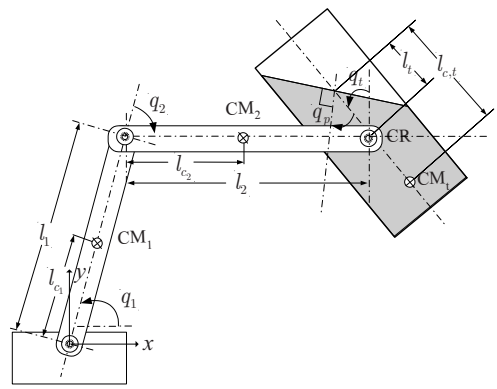


Figure 1. Liquid transfer application [15]

As it has been illustrated in [15], this control scheme achieves sufficiently accurate positioning of the tank with simultaneous liquid sloshing suppression, in the ideal case when the physical parameters of the liquid, and consequently the parameters of the pendulum are known accurately. However, in most practical applications these parameters are subject to uncertainty, fact that deteriorates the closed-loop performance.

In the following sections, the robustness of the two stage control scheme, introduced in [15], is studied with respect to uncertainties on the physical parameters of the liquid. The range of uncertainties that can be well tolerated without critical deterioration of the closed-loop performance is determined using simulation results. The closed-loop performance is evaluated based on the maximum amplitude of liquid's oscillations and the maximum deviation of the tank's position from the desired trajectory. The simulation results illustrate that the proposed control scheme is sufficiently robust with respect to liquid sloshing suppression. However, positioning of the tank appears to be quite more sensitive with respect to parameter uncertainties. In order to improve robustness, we propose the incorporation in the control scheme of additional integral terms, which aim to improve closed-loop performance with respect to command following for the tank's positioning.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 154–161.

Section II presents in short the robotic liquid transfer structure and the considerations used to derive the corresponding nonlinear model. Moreover, Section II discusses the reasons that may cause parametric uncertainty in this model and presents the considered modeling of the uncertainty. Section III presents in short the control design scheme proposed in [15]. Section IV studies the robust performance of this control scheme. More specifically, it presents the configuration of the simulations used to study robustness, the performance of the nominal closed-loop system and the derived robust analysis results. Finally, Section V presents the proposed enhancement of the control design scheme.

II. Liquid Transfer Structure

A. Nonlinear Model of the Liquid Transfer Structure

The control design scheme proposed in [15] concerns the robotic liquid transfer structure presented in Figure 1, where an articulated planar manipulator carries a tank filled with a liquid. In order to simplify modeling, the dynamics of the liquid transfer structure are approximated by the dynamics of the robotic structure presented in Figure 2, where the tank that contains the liquid is considered to be represented by a single link, while the liquid's motion is represented by a pendulum, which rotates freely around an axis that is perpendicular to the structure's plane of motion and passes through the upper point of the link that represents the tank. The motion of the link that represents the tank is actuated through an appropriate torque applied by the manipulator's end-effector.

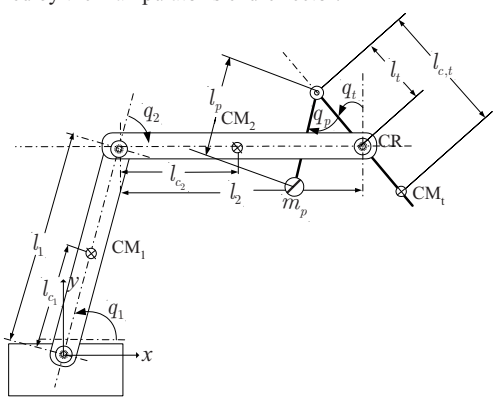


Figure 2. Representation of liquid's motion with a pendulum [15]

The performance variables of the liquid transfer structure are the generalized variables q_1 and q_2 of the robotic manipulator, the angle q_t of the tank's rotation with respect to the perpendicular axis and the rotation q_p of the pendulum with respect to the tank, which denotes the angle between the liquid's free surface with the bottom of the tank. The control variables of the structure are the torques $u_1(t)$, $u_2(t)$ and $u_3(t)$ that actuate the first manipulator joint, the second manipulator joint and the tank's rotation, respectively.

The Euler-Lagrange dynamic equations of the simplified robotic structure are given by:

$$D(q(t))\ddot{q}(t)+C(q(t),\dot{q}(t))\dot{q}(t)+G(q(t))=u(t) \qquad (1)$$

where

$$q=\begin{bmatrix} q_1 & q_2 & q_t & q_p \end{bmatrix}^{\mathrm{T}},\ u(t)=\begin{bmatrix} u_1(t) & u_2(t) & u_3(t) & 0 \end{bmatrix}^{\mathrm{T}} \qquad (2)$$

The elements of the matrices $D(q(t))=\left[d_{ij}(q(t))\right]$, $C(q(t),\dot{q}(t))=\left[c_{ij}(q(t),\dot{q}(t))\right]$, $G(q(t))=\left[G_i(q(t))\right]$, $i,j=1,\ldots,4$ are nonlinear functions of the structure's generalized variables and their respective velocities. Note that $D(q(t))$ is a symmetric positive definite matrix. The elements of these matrices are given in [15].

The parameters of model (1), which are given in Table 1, depend on the characteristics of the robot's joints, the tank and the pendulum. The mass m_t and the inertia I_t of the link that represents the tank are equal to the corresponding parameters of the tank. The mass m_p of the pendulum is equal to the mass of the liquid. The equivalent length l_p of the pendulum is determined based on the natural frequency given by the perfect fluid theory [1]. Moreover, the pendulum-type sloshing model considers a torque applied to the pendulum due to the viscosity of the liquid and the friction between the liquid and the walls of the tank and which is equal to $-cl_p^2\cos^2(q_p+q_t)\dot{q}_p$, where c an equivalent coefficient of viscosity [1].

NOMENCLATURE

Parameter	Nominal Value	Physical Meaning
m_1	10[kg]	Mass of the 1st link [17]
m_2	10[kg]	Mass of the 2nd link [17]
l_1	1[m]	Length of the 1st link [17]
l_2	1[m]	Length of the 2nd link [17]
I_1	$2[\mathrm{kg}\cdot\mathrm{m}^2]$	Moment of inertia of the 1st link [17]
I_2	$2[\mathrm{kg}\cdot\mathrm{m}^2]$	Moment of inertia of the 2nd link [17]
l_{c1}	0.5[m]	Distance of the 1st link's center of mass from the 1st joint [17]
l_{c2}	0.5[m]	Distance of the 2nd link's center of mass from the 2nd joint [17]
h_t	0.17[m]	Tank's height
m_t	$1.68[\mathrm{kg}]$	Tank's mass
$l_{t,n}$	$0.033[\mathrm{m}]$	Nominal distance between the tank's center of rotation (CR) and the free surface of the liquid
$l_{ct,n}$	$0.113[\mathrm{m}]$	Nominal distance of the tank's center of mass (CM) from the free surface of the liquid
I_t	$7.193\cdot10^{-3}[\mathrm{kg}\cdot\mathrm{m}^2]$	Tank's moment of inertia

$m_{p,n}$	$2.744[\mathrm{kg}]$	Nominal liquid's mass [1]
$l_{p,n}$	$0.044[\mathrm{m}]$	Nominal pendulum's length [1]
ρ_n	$1000\left[\mathrm{kg/m^3}\right]$	Nominal density of the liquid (water)
c_n	$1.88[\mathrm{N\cdot second/m}]$	Nominal liquid's equivalent coefficient of viscosity [1]
$h_{s,n}$	$0.14[\mathrm{m}]$	Nominal level of the liquid in the tank [1]
g	$9.81\left[\mathrm{m/second^2}\right]$	Gravity acceleration

Table 1. Parameters of the liquid transfer structure

B. Parametric uncertainty

As already mentioned in the previous subsection, the liquid's motion is approximated by a pendulum-type model. The parameters of this model depend on the physical parameters of the liquid, namely the liquid's density ρ, the liquid's level h_s within the tank and finally, the liquid's equivalent coefficient of viscosity c. In general, nominal values for these parameters are usually available for control design. However, their true values may deviate from the nominal ones.

More specifically, deviations of the liquid's level h_s within the tank from its nominal value, may appear due to erroneous measurements, accidental liquid outpouring, as well as due to intentional liquid outpouring during a pouring process.

Deviations on the liquid's density ρ and equivalent coefficient of viscosity c from their nominal values, may occur due to incidental variations of the liquid's characteristics in previous steps of the productive procedure. Moreover, deviations on these parameters may occur when the same liquid transfer structure is used to carry several liquids with different physical properties.

Deviations which are due to incidental factors are usually relatively small and should be well tolerated by a control design scheme. On the other hand, deviations which are caused intentionally, as results of specific tasks of the productive procedure, may have significant amplitudes, that cannot be always treated effectively by a control scheme designed for the nominal values of the liquid's parameters. Although such variations may be known to the designer, it is beneficial for practical reasons to apply a control scheme which is able to tolerate a sufficient range of parameters' deviations from their nominal values. This would reduce the need for controller retuning whenever the liquid's parameters values have to change according to the needs of the productive procedure.

Whatever the cause, parameter value deviations from their nominal values may be modeled as additive uncertainties on the liquid's parameters:

$$h_s = h_{s,n} + p_h \quad , \quad \rho = \rho_n + p_\rho \quad , \quad \tilde{c} = \tilde{c}_n + p_{\tilde{c}} \qquad (3)$$

where h_s, ρ and $\tilde{c} = c/\rho$ are the true values of the liquid's parameters, $h_{s,n}$, ρ_n and $\tilde{c}_n = c_n/\rho_n$ are the corresponding nominal values, while p_h, p_ρ and $p_{\tilde{c}}$ denote the additive uncertainties. Note that the uncertainties are modeled with respect to the parameter $\tilde{c} = c/\rho$, which will be called from now on normalized viscosity coefficient. This was motivated by the dynamic equations of the pendulum given in [1] and [5]. Since the liquid's mass m_p is analogous to the density ρ, it follows from [1] and [5], that dividing both sides of the pendulum's dynamic equations by ρ and setting $\tilde{c} = c/\rho$, then the density does no longer appear as an uncertain parameter of the pendulum's dynamics. Note, however, that the liquid's density remains as an uncertain parameter in the Euler Lagrange dynamic equations (1), since these equations incorporate parameters which vary with h_s and ρ.

More specifically, the liquid's mass m_p is given by

$$m_p = \rho h_s S \qquad (4)$$

where S the area of the tank's bottom. The equivalent length l_p of the pendulum is determined based on the natural frequency given by the perfect fluid theory [5], [7]. Assuming the dimension of the sloshing mode to be equal to one, the natural frequency f_s is related to the liquid level h_s according to the relation ([5], [7]) $f_s = \frac{1}{2\pi}\sqrt{\frac{g\pi}{R}\tanh\left(\frac{\pi h_s}{R}\right)}$, where R the distance between the walls of the tank. Then the pendulum's length l_p is given by:

$$l_p = \frac{g}{4\pi^2 f_s^2} = \frac{R}{\pi \tanh\left(\pi h_s / R\right)} \qquad (5)$$

which implies that l_p varies with the liquid's level h_s. Finally, the distances l_t and l_{ct} between the free surface of the liquid and the tank's center of rotation and center of mass, respectively, are given by:

$$l_t = l_{t,n} + h_s - h_{s,n}, \quad l_{ct} = l_{ct,n} + h_s - h_{s,n} \qquad (6)$$

where $l_{t,n}$ and $l_{ct,n}$ the corresponding nominal values given in Table 1. The dynamic equations (1) derived in [15] hold under the assumption that $l_t \geq 0$. Hence, these equations may be used to study variations of h_s that satisfy the condition:

$$h_s \geq h_{s,n} - l_{t,n} \qquad (7)$$

III. Two Stage Control Design Scheme

In [15] a two stage control design scheme has been proposed to control the liquid transfer structure presented in the previous section. This design scheme combines a partial inverse dynamics controller with a PID controller. Inverse dynamics controllers are designed to achieve linearization and input/output decoupling of a nonlinear system's dynamics.

However, in practical applications, measurements of the pendulum position and velocity are usually not available, thus preventing the design of a fully linearizing inverse dynamic controller for system (1). To deal with this problem, the following partial inverse dynamics controller has been proposed in [15], which would achieve linearization and input/output decoupling of the liquid's transfer structure dynamics, as well as asymptotic tracking for the generalized variables of the two joints and the tank, in the ideal case where no liquid vibrations take place:

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \bar{C}(\bar{q},\dot{\bar{q}})\dot{\bar{q}} + \bar{G}(\bar{q}) + \bar{D}(\bar{q}) \begin{bmatrix} \ddot{q}_{d,1} + k_1(\dot{q}_{d,1} - \dot{q}_1) + k_2(q_{d,1} - q_1) \\ \ddot{q}_{d,2} + k_1(\dot{q}_{d,2} - \dot{q}_2) + k_2(q_{d,2} - q_2) \\ -k_1\dot{q}_t - k_2 q_t + k_2 w_3 \end{bmatrix} \quad (8)$$

where $\bar{q} = [q_1 \quad q_2 \quad q_t]^{\mathrm{T}}$, $w_3(t)$ is an auxiliary input variable, and

$$\bar{D}(\bar{q}) = [d_{i,j}([\bar{q}^{\mathrm{T}} \quad 0]^{\mathrm{T}})], \; i,j = 1,2,3 \quad (9a)$$

$$\bar{C}(\bar{q},\dot{\bar{q}}) = [c_{i,j}([\bar{q}^{\mathrm{T}} \quad 0]^{\mathrm{T}},[\dot{\bar{q}}^{\mathrm{T}} \quad 0]^{\mathrm{T}})], \; i,j = 1,2,3 \quad (9b)$$

$$\bar{G}(\bar{q}) = [G_i([\bar{q}^{\mathrm{T}} \quad 0]^{\mathrm{T}})], \; i = 1,2,3 \quad (9c)$$

while $q_{d,1}(t)$ and $q_{d,2}(t)$ denote the desired trajectories for the first and the second manipulator's joint variables, respectively. The desired trajectory for the tank's rotation angle is identically equal to zero. The controller parameters k_1 and k_2 are appropriately selected so as to achieve closed-loop stability and sufficiently small settling time. Note that the proposed partial inverse dynamics controller exploits measurements of the manipulator links and the container position and velocity variables, but it does not require measurements of the pendulum position and velocity.

However, the aforementioned controller is inadequate with respect to sloshing suppression, while its performance with respect to the tank's positioning would also deteriorate under the presence of significant liquid vibrations. To deal with this problem, an additional PID controller has been proposed in [15], in order to achieve sloshing suppression. This PID controller is designed to drive the auxiliary input variable $w_3(t)$, considering as output the tank's angle of rotation $q_t(t)$, i.e. [15]:

$$w_3(t) = -f_1 q_t(t) - f_2 \dot{q}_t(t) - f_3 \int_0^t q_t(\tau)\, d\tau \quad (10)$$

where $f_i, i = 1,2,3$ are the controller parameters, which are tuned with the use of a metaheuristic search algorithm [16]. The values of the PID controller parameters have been determined in [15] to be equal to:

$$f_1 = 74.7907, \; f_2 = 2.5986, \; f_3 = 21.4256 \quad (11)$$

Note that the metaheuristic algorithm utilizes simulation results of the whole plant's dynamics, where the parameters of the pendulum are considered to be equal to their nominal values given in Table 1.

The performance of the closed-loop system with respect to liquid sloshing suppression may be evaluated by the maximum absolute value $q_{p\max}$ of the pendulum's angle $q_p(t)$, i.e. $q_{p\max} = \max_{t\geq 0}\{|q_p(t)|\}$. The maximum value $q_{p\max}$ is a critical performance variable, since large values of $q_{p\max}$ may result in undesired liquid overflow and excessive cooling of the liquid.

The performance of the closed-loop system with respect to positioning of the tank may be evaluated by the maximum deviations of the $x_e(t)$ and $y_e(t)$ coordinates of the end-effector's center of rotation CR from the corresponding desired trajectories $x_d(t)$ and $y_d(t)$. The performance requirements in pouring applications usually require motion of the tank between an initial and a final position. The final position has to be achieved with significant accuracy, since it is determined by the position of the mold which has to filled with the liquid. However, during transition larger errors with respect to trajectory following may be well tolerated. For this reason, the performance with respect to tank's positioning uses the following four measures of evaluation:

$$e_{xt} = \max_{t\in[0,t_p]}\{|x_e(t) - x_d(t)|\}, e_{yt} = \max_{t\in[0,t_p]}\{|y_e(t) - y_d(t)|\} \quad (12a)$$

$$e_{xs} = \max_{t>t_p}\{|x_e(t) - x_d(t)|\}, \; e_{ys} = \max_{t>t_p}\{|y_e(t) - y_d(t)|\} \quad (12b)$$

Note that $e_{x,t}$ and $e_{y,t}$ are used to evaluate performance with respect to command following during the transition stage, while $e_{x,s}$ and $e_{y,s}$ are used to evaluate performance with respect to tank's final positioning. The selection of t_p introduces an additional performance requirement, since it determines the time interval within which the desired final positioning has to achieved.

IV. Robust Performance Analysis

A. Simulation Configuration – Nominal Performance

The performance of the control design scheme, presented in the previous section, has been illustrated in [15] through simulation results for the nominal case, where the pendulum model parameters are equal to their nominal values given in Table 1. As already mentioned, these are also the values used for controller tuning.

The simulations are derived assuming that the desired trajectories of the robotic manipulator's generalized variables q_1 and q_2 are determined so as to move the point CR, at

which the manipulator grasps the tank, along a straight path starting from the point with coordinates (1,1) at $t = 0$ and reaching the point with coordinates (1.27,1.27) at $t = 2$. As indicated in [15], this motion is achieved by selecting the desired trajectories of the generalized variables q_1 and q_2 equal to the solutions of the following inverse kinematics equations:

$$\begin{aligned} l_1 \cos(q_{d,1}(t)) + l_2 \cos(q_{d,1}(t) + q_{d,2}(t)) &= x_d(t) \\ l_1 \sin(q_{d,1}(t)) + l_2 \sin(q_{d,1}(t) + q_{d,2}(t)) &= y_d(t) \end{aligned} \quad (13)$$

where the desired trajectories $x_d(t)$ and $y_d(t)$ are determined by the equations:

$$x_d(t) = 1 + 0.5\,(1.8\cos(\pi/4) - 1)r(t) \quad (14a)$$

$$y_d(t) = 1 + 0.5(1.8\sin(\pi/4) - 1)r(t) \quad (14b)$$

$$r(t) = (3t^5 - 15t^4 + 20t^3)/8 \quad (14c)$$

The initial conditions for the state variables of system (1) are considered to be:

$$\begin{gathered} q_1(0\text{-}) = \pi/2[\text{rad}],\ q_2(0\text{-}) = -\pi/2[\text{rad}], \\ q_t(0\text{-}) = 0[\text{rad}],\ q_p(0\text{-}) = 0[\text{rad}], \\ \dot{q}_1(0\text{-}) = 0[\text{rad/second}],\ \dot{q}_2(0\text{-}) = 0\ [\text{rad/second}], \\ \dot{q}_t(0\text{-}) = 0[\text{rad/second}],\ \dot{q}_p(0\text{-}) = 0[\text{rad/second}]. \end{gathered}$$

Figure 3, which presents the closed-loop variation of the pendulum's angle, illustrates the nominal closed-loop performance with respect to liquid sloshing. It is obvious from Figure 3 that the oscillations of the liquid are damped sufficiently fast, while being kept within sufficiently small bounds during the transient stage. More specifically, the maximum value of q_p is equal to

$$q_{p\max,n} = 0.047[\text{rad}] \quad (15)$$

Figure 4 illustrates the closed-loop performance with respect to tank's positioning. As it follows from Figure 4, the proposed control scheme achieves perfect positioning of the tank. Indeed, selecting $t_p = 3.5[\text{seconds}]$, the maximum values of the corresponding deviations of the closed-loop variables $x_e(t)$ and $y_e(t)$ from the desired trajectories $x_d(t)$ and $y_d(t)$, respectively, are given by:

$$e_{xt,n} = 7.758 \cdot 10^{-4}, e_{yt,n} = 2.949 \cdot 10^{-4} \quad (16a)$$

$$e_{xs,n} = 1.185 \cdot 10^{-5},\ e_{ys,n} = 5.167 \cdot 10^{-6} \quad (16b)$$

Note that the selection of t_p leaves a time window of $1.5[\text{second}]$ for the tank to settle to its final position, which is considered sufficient, taking into account that the desired trajectory assumes that the final position is reached within $2[\text{seconds}]$.

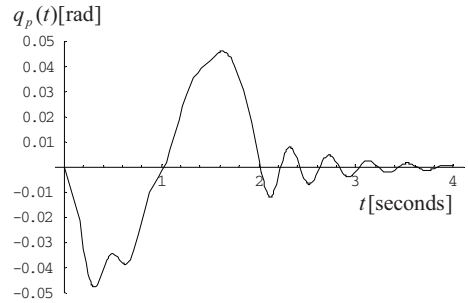


Figure 3. Closed-loop values for $q_p(t)$ - Nominal performance

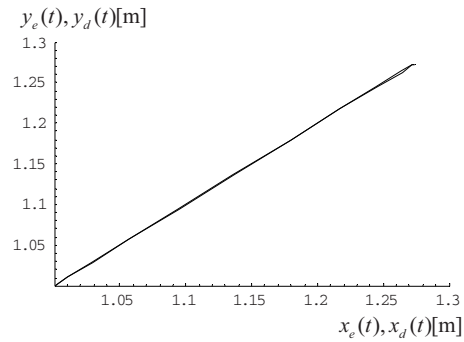


Figure 4. Desired and closed-loop coordinates of the point CR (visually identical) – Nominal performance

B. Robust Analysis

In this section we study the robust performance of the control design scheme presented in Section III, with respect to the parametric uncertainty described in (3). The evaluation of robustness is performed using simulations for several values of the uncertain parameters. The simulation results are used to compute the maximum value $q_{p\max}$ of the variable q_p, as well as the command following errors e_{xt}, e_{yt}, e_{xs} and e_{ys}, determined in (12). Uncertainties are considered to be well tolerated provided that the following conditions hold:

$$q_{p\max} \le 1.2 q_{p\max,n} \quad (17a)$$

$$e_{xt} \le 0.03[\text{m}],\ e_{yt} \le 0.03[\text{m}] \quad (17b)$$

$$e_{xs} \le 0.02[\text{m}],\ e_{ys} \le 0.02[\text{m}] \quad (17c)$$

Condition (17a) implies that the maximum amplitude of liquid's oscillations does not deteriorate more than 20% due to uncertainties. Satisfaction of this condition also guarantees that no undesired liquid outpouring occurs. Condition (17b) implies that at each time instant $t \le 3.5[\text{seconds}]$ the point CR lies within a square centered on the corresponding point of the

desired trajectory, whose side is equal to 0.06[m]. Condition (17c) implies that at each time instant $t > 3.5$[seconds] the point CR lies within a square centered on the desired final position, whose side is equal to 0.04[m].

Simulation results were performed for a sufficiently dense web of points within the following areas of uncertainty:

$$p_h \in [-0.03, 0.01][\text{m}] \quad , \quad p_\rho \in [-500, 700][\text{kg/m}^3] \qquad (18a)$$

$$p_{\tilde{c}} \in [-1.25 \cdot 10^{-3}, 48.88 \cdot 10^{-3}][\text{m}^3/\text{second}] \qquad (18b)$$

The range of uncertainty p_h is determined taking into account condition (7), as well the height of the tank, given in Table 1. The range of uncertainty p_ρ is considered to give a reasonable variation, taking into account that the control scheme is designed for nominal value equal to 1000[kg/m^3]. The range of uncertainty $p_{\tilde{c}}$ is selected very wide, so as to cover the majority of practical needs.

It is important to note at this point, that as indicated in [3] and [8], the parameter c is an equivalent viscosity coefficient that may vary with the liquid's level h_s. Moreover, some combinations of density ρ and viscosity coefficient c may not be met in practice. Hence, there may exists points within the domain of uncertainty given by (18), which may not have practical value. However, since the dependence between the liquid's physical parameters is too complex and partially unknown, robust analysis is performed considering these parameters to be independent.

According to simulation results, the closed-loop performance is robust with respect to sloshing suppression, since $q_{p\max}$ remains smaller than $1.146 q_{p\max,n}$ for all values of uncertain parameters within the range of uncertainty determined in (18).

Also according to simulation results, although deviations of the normalized coefficient $\tilde{c}$ may have effect on the amplitude of liquid oscillations, they affect only slightly the closed-loop performance with respect to positioning of the tank. For this reason, robust analysis results are presented in the following with respect to variations of h_s and ρ. Figure 5 presents the area of uncertain parameters where robust performance is achieved with respect to positioning of the tank, that is the area within which conditions (17b) and (17c) are satisfied. Figures 6 and 7 illustrate the robustness of the closed-loop performance for liquid's parameters values given by

$$\begin{aligned} &h_s = 0.11[\text{m}] \ (p_h = -0.03[\text{m}]) \\ &\rho = 1525[\text{kg/m}^3] (p_\rho = 525[\text{kg/m}^3]), \\ &\tilde{c} = 0.63 \cdot 10^{-3}[\text{m}^3/\text{second}] \ (p_{\tilde{c}} = -1.38 \cdot 10^{-3}[\text{m}^3/\text{second}]) \end{aligned} \qquad (19)$$

Note that these values of h_s and ρ correspond to the upper left vertex of the area represented in Figure 5.

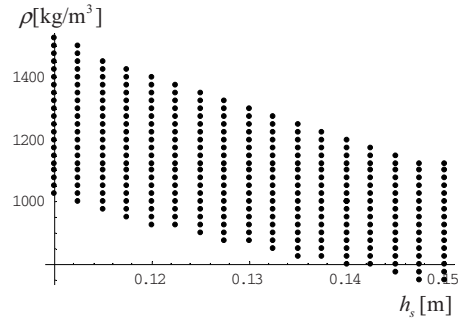


Figure 5. Results of robust performance with respect to tank's positioning

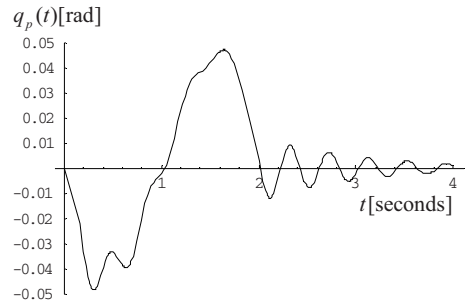


Figure 6. Closed-loop values for $q_p(t)$ - Uncertain parameters given by (19)

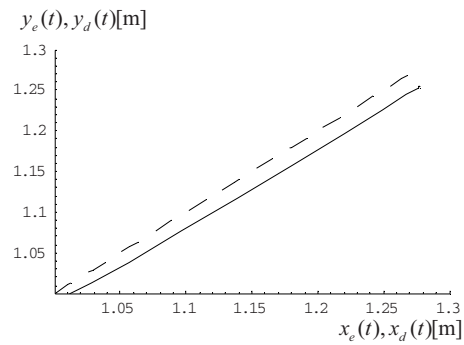


Figure 7. Desired (dashed line) and closed-loop (solid line) coordinates of the point CR - Uncertain parameters given by (19)

V. Integral Control for Tank's Positioning

As illustrated in Figure 5, the range of uncertainty where robustness with respect to tank's positioning is achieved is not sufficiently large. In order to extend this range, the following modification of the partial inverse dynamics controller given in (8) is proposed:

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \bar{C}(\bar{q}, \dot{\bar{q}})\dot{\bar{q}} + \bar{G}(\bar{q}) + \bar{D}(\bar{q}) \times$$

$$\begin{bmatrix} \ddot{q}_{d,1} + k_1(\dot{q}_{d,1} - \dot{q}_1) + k_2(q_{d,1} - q_1) + k_2 \int_0^t \left[q_{d,1}(\tau) - q_1(\tau) \right] d\tau \\ \ddot{q}_{d,2} + k_1(\dot{q}_{d,2} - \dot{q}_2) + k_2(q_{d,2} - q_2) + k_2 \int_0^t \left[q_{d,2}(\tau) - q_2(\tau) \right] d\tau \\ -k_1 \dot{q}_t - k_2 q_t + k_2 w_3 \end{bmatrix}$$

As it follows from the previous equation, the proposed modification consists in the incorporation of two integral error feedback terms, with gains equal to k_2, which aim to reduce command following error with respect to the manipulator joint variables q_1 and q_2.

Simulation results indicate that the proposed modification improves the robustness characteristics of the control scheme with respect to both sloshing suppression and tank's positioning. Indeed, $q_{p\max}$ remains smaller than $1.087 q_{p\max,n}$ for all values of uncertain parameters within the range of uncertainty determined in (18). Moreover, Figure 8 illustrates the enlargement of the uncertain parameter area where conditions (17b) and (17c) are satisfied, in comparison with the corresponding area of Figure 5. Finally, for comparison reasons, Figures 9 and 10 illustrate the corresponding closed-loop performance for liquid's parameter values given by (19).

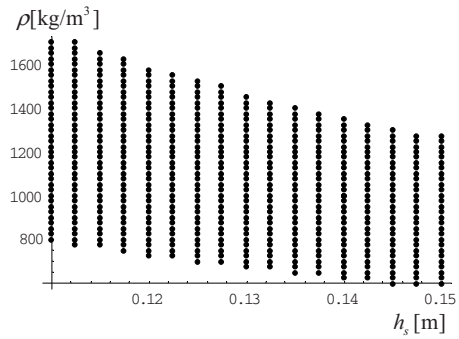


Figure 8. Results of robust performance with respect to tank's positioning – Modified control scheme

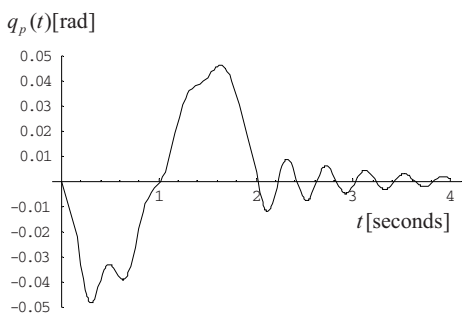


Figure 9. Closed-loop values for $q_p(t)$ - Uncertain parameters given by (19) - – Modified control scheme

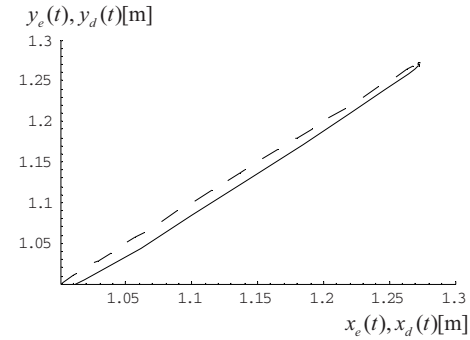


Figure 10. Desired (dashed line) and closed-loop (solid line) coordinates of the point CR - Uncertain parameters given by (19) – Modified control scheme

VI. Conclusions

In the present work, the robust performance of a two stage control design scheme for sloshing suppression and container's positioning during liquid transfer using robotic manipulators, has been studied with respect to parametric uncertainty. The control scheme, that was introduced in [15], combines a partial inverse dynamics controller with a heuristically tuned PID controller. The closed-loop performance is evaluated with respect to the amplitude of liquid's oscillations and the command following error in the tank's motion. It has been illustrated using simulation results, that the two stage robot control scheme is robust with respect to uncertainties on the physical parameters of the liquid. The corresponding range of uncertainty, where robustness is achieved, has been determined. Moreover, a modification of the control scheme has been proposed in order to extend the range of uncertainty that can be well tolerated with respect to tank's positioning. The present results increase the practical value of the proposed control scheme by reducing the need for controller retuning, provided that the liquid parameters' variations remain within appropriate bounds.

Acknowledgment

The present work is co-financed by the Hellenic Ministry of Education and Religious Affairs' and the ESF of the European Union within the framework of the "Operational Programme for Education and Initial Vocational Training" (Operation "Archimedes-II").

References

[1] K. Yano, S. Higashikawa and K. Terashima, "Motion control of liquid container considering an inclined transfer path," *Control Engineering Practice*, vol. 10, pp. 465-472, 2002.

[2] J. Feddema, C. Dohrmann, G. Parker, R. Robinett, V. Romero and D. Schmitt, "Robotically controlled slosh-free motion of an open container of liquid," in *Proc. of the 1996 IEEE International Conf. on Robotics and Automation*, Mineapolis, Minnesota, 1996, pp. 596-602.

[3] K. Terashima, M, Hamaguchi and K. Yamaura, "Modeling and input shaping control of liquid vibration for an automated pouring system," in *35th Conf. on Decision Control*, Kobe, Japan, 1996, pp. 4844-4850.

[4] J. Feddema, C. Dohrmann, G. Parker, R. Robinett, V. Romero and D. Schmitt, "Control for slosh-free motion of an open container," *IEEE Control Systems Magazine*, vol. 17, no. 1, pp. 29-36, 1997

[5] K. Yano and K. Terashima, "Robust liquid container transfer control for complete sloshing suppression," *IEEE Transactions on Control Systems Technology*, vol. 9, no. 3, pp. 483-493, 2001.

[6] K. Yano, S. Higashikawa and K. Terashima, "Liquid container transfer control on 3D transfer path by hybrid shaped approach," *2001 IEEE Int. Conf. on Control Applications*, 2001, Mexico City, pp. 1168-1173.

[7] K. Yano, T. Toda and K. Terashima, "Sloshing suppression control of automatic pouring robot by hybrid shape approach,", *40th IEEE Conf. on Decision and Control*, Orlando, Florida, USA, 2001, pp. 1328-1333.

[8] K. Terashima and K. Yano, "Sloshing analysis and suppression control of tilting-type automatic pouring machine," *Control Engineering Practice*, vol. 9, pp. 607-620, 2001

[9] H. Sira-Ramirez, "A flatness based generalized PI control approach to liquid sloshing regulation in a moving container," in *Proc. of the American Control Conf.*, Anchorage, USA, 2002, pp. 2909-2914.

[10] S. Kimura, M. Hamaguchi and T. Taniguchi, "Damping control of liquid container by a carrier with dual swing type active vibration reducer," in *Proc. of the 41st SICE Annual Conf.*, 2002, pp. 2385-2388.

[11] Y. Noda, K. Yano and K. Terashima, "Tracking to moving object and sloshing suppression control using time varying filter gain in liquid container transfer," *2003 SICE Annual Conf.*, Fukui, Japan, 2003, pp. 2283-2288.

[12] M. Hamaguchi, K. Terashima, H. Nomura, "Optimal control of liquid container transfer for several performance specifications," *Journal of Advanced Automation Technology*, vol. 6, pp. 353-360, 1994.

[13] J. Feddema, C. Dohrmann, G. Parker, R. Robinett, V. Romero and D. Schmitt, "A comparison of maneuver optimization and input shaping filters for robotically controlled slosh-free motion of an open container of liquid", *American Control Conf.*, Albuquerque, New Mexico, 1997, pp. 1345-1349.

[14] M.P. Tzamtzi, F.N. Koumboulis, N.D. Kouvakas, G.E. Panagiotakis, "A Simulated Annealing Controller for Sloshing Suppression in Liquid Transfer", *14th Mediterranean Conf. on Control and Automation*, Ancona, Italy, 2006.

[15] M.P. Tzamtzi, F.N. Koumboulis, N.D. Kouvakas, "A two stage robot control for liquid transfer", *12th IEEE Conf. on Emerging Technologies and Factory Automation (ETFA 2007),* Greece, pp. 1324-1333, 2007.

[16] F.N. Koumboulis, M.P. Tzamtzi, "A metaheuristic approach for controller design of multivariable processes", *12th IEEE Conf. on Emerging Technologies and Factory Automation (ETFA 2007),* Patras, Greece, pp. 1429-1432, 2007.

[17] B. Yao, W.B. Gao, S.P. Chan and M. Cheng, "VSC coordinated control of two manipulator arms in the presence of environmental constraints", *IEEE Trans. Automat. Contr.,* vol. 37, pp. 1806-1812, 1992.

[18] C. Canudas de Wit, B. Siciliano, G. Bastin, *Theory of Robot Control*, Springer- Verlag, London, 1996.

A Process Application of Step-Wise Safe Switching Control

F. N. Koumboulis
Department of Automation
Halkis Institute of Technology
34400 Psahna, Evia, Greece
koumboulis@teihal.gr

M. P. Tzamtzi
Department of Automation
Halkis Institute of Technology
34400 Psahna, Evia, Greece
tzamtzi@teihal.gr

Abstract- **In the present work we study the application to a characteristic nonlinear process of the Step-Wise Safe Switching (SWSS) control approach for SISO plants. The process nonlinear description is considered to be unknown for control design, while its behavior is approximated by a set of SISO linear models derived through identification around corresponding operating points. The performance of the supervisory scheme is illustrated using simulation results, which are also compared with corresponding results of other approaches.**

I. Introduction

Supervisory switching control for nonlinear processes (see for example [1]-[8]) comprises, in general, a set of field controllers and a supervisory controller. Each field controller achieves the required closed-loop performance provided that the plant's input/output (I/O) trajectories remain within a limited range of operation. The supervisory controller implements switching between the appropriate field controllers, as the plant's I/O trajectories move between different areas of operation.

A switching approach, called Step-Wise Safe Switching (SWSS), has been first established in [4] for the case of single input-single output (SISO) systems, with unknown description. The plant is modeled using a switching multi-linear model, comprising a set of SISO linear systems, which are derived through identification around a sufficiently dense set of nominal operating points. A significant characteristic of SWSS is the application of "common" controllers that achieve the required performance simultaneously for more than one adjacent nominal operating points. Moreover, controller switching is allowed only when the process is close to steady state. This requirement avoids undesirable effects that may come from switching while moving e.g. instability. Thus, the motion between any two different operating points is performed by moving in a step-wise manner between operating areas of an appropriately selected sequence of adjacent nominal operating points. The operating areas are determined with the use of experimental process data. In [4], the performance of SWSS has been studied for the case of the first-order nonlinear process of a CSTR reactor.

In the present work we study the application to a characteristic nonlinear process with second-order dynamics of the SWSS control approach for SISO plants. The process nonlinear description is considered to be unknown for control design, while its behavior is approximated by a set of linear models derived through identification around corresponding operating points. The performance of the SWSS scheme, which is studied using simulation results, seems to be advantageous when compared to a standard switching algorithm, as well as when compared to an approach seeking for "common" (robust) controllers for more than two operating points. The performance of SWSS for plants with higher order dynamics will be studied in future works.

II. Step-Wise Safe Switching

In the present section, the SWSS algorithm for SISO systems introduced in [4], is presented in short (see also [7]).

Consider a SISO plant with discrete-time description, where $y(k)$ denotes the plant's output and $u(k)$ denotes the plant's input. Let $L=\{\ell_1,\ell_2,\ldots,\ell_\mu\}$ denote a set of nominal operating points of the plant (points of the input-output space where the plant may settle at steady state), where each $\ell_i, i=1,\ldots,\mu$ is denoted as $\ell_i=[Y_i,U_i]$, with Y_i and U_i denoting the corresponding nominal output and input values. Consider also that the plant's description is approximated by the following set of linear models, which are determined through identification around the nominal operating points:

$$S_i:\ A_i(q^{-1})\Delta_i y(k)=B_i(q^{-1})\Delta_i u(k-d_i)+C_i(q^{-1})\varepsilon_i(k)$$

where $\Delta_i y(k)=y(k)-Y_i$ and $\Delta_i u(k)=u(k)-U_i$ denote perturbations of the output and input variables around ℓ_i and $\varepsilon_i(k)$ denotes the unmodeled error or disturbance in S_i. The nonnegative integer d_i denotes the delay of the model S_i. The operators $A_i(q^{-1})$, $B_i(q^{-1})$ and $C_i(q^{-1})$ are polynomials of the delay operator q^{-1} with real coefficients.

For each nominal operating point ℓ_i we determine a pair of operating areas, named *target* (O_i) and *tolerance* ($\bar{O}_i$) operating areas [4], which constitute experimental approximations of the neighborhood of validity of each local linear model S_i. The target operating area is determined as a rectangle in the (U,Y)-space, according to the following rule: For each step transition between an initial operating point $\rho_s=[Y_s,U_s]$ within O_i to a final operating point

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 162–167.

$\rho_f = [Y_f, U_f]$ also within O_i, the percentage of deviation between the responses of the nonlinear system and the corresponding linearized system S_i remain smaller than a threshold value $\varepsilon_{\text{target}}$. The tolerance operating area is determined in a similar way using a threshold value $\varepsilon_{\text{tol}} > \varepsilon_{\text{target}}$. Note that the target and the tolerance operating areas are determined using exclusively experimental data.

The nominal operating points are selected dense enough to satisfy the following requirements [4]: $O_i \cap O_{i+1} \neq \varnothing$, $i = 1,\ldots,\mu-1$, $O_1 \cup O_2 \subset \overline{O}_1$, $O_{\mu-1} \cup O_\mu \subset \overline{O}_\mu$, $O_{i-1} \cup O_i \subset \overline{O}_i$, $O_i \cup O_{i+1} \subset \overline{O}_i$, $i = 2,\ldots,\mu-1$. The above conditions constitute an experimental formulation of the *dense web principle* [4], according to which the linear models S_i, $i = 1,\ldots,\mu$ describe satisfactorily the process behavior.

Finally, consider that for each pair (ℓ_i, ℓ_{i+1}) of adjacent operating points, there exists a "common" controller $C_{i,i+1}$ that satisfies a set of desired design requirements simultaneously for both linear models S_i and S_{i+1}. Then, the SWSS algorithm is summarized in the following steps [4]:

Step-Wise Safe Switching Algorithm

Step 1: Apply a safe controller until the output variable is relaxed (i.e. reaches a steady state).
Step 2: Set $\lambda = 1$.
Step 3: Read the present operating point, let ρ_λ.
Step 4: Read the desired operating point $\rho_{\lambda+1}$.
Step 5: Choose a pair of adjacent target operating areas $(O_{\eta_\lambda}, O_{\eta_{\lambda+1}})$, such that ρ_λ lies within O_{η_λ} and $\rho_{\lambda+1}$ lies within $O_{\eta_{\lambda+1}}$, else return to Step 1.
Step 6: Switch to controller $C_{\eta_\lambda,\eta_{\lambda+1}}$ and force the closed loop system from ρ_λ to $\rho_{\lambda+1}$.
Step 7: While the I/O values of the process remain within the areas $\overline{O}_{\eta_\lambda} \cup \overline{O}_{\eta_{\lambda+1}}$ wait for a time period $t \le t_{\max}$ or till the system approaches a steady state namely an operating point $\rho^*_{\lambda+1}$, else return to Step 1.
Step 8: Set $\lambda = \lambda + 1$ and return to Step 3.

III. Process Description

Consider the process of a double effect evaporator with short-tube vertical calandria-type units, described in [10] and [11]. The process input u is the flow rate of a saturated steam, that heats the feed solution pumped to the first effect. The solution produced by the first effect is fed to the second effect, where it is in turn heated by the vapor flow produced by the first effect. The solution concentration produced by the second effect is the process output y. The double effect evaporator process is a stable nonlinear process. Its equilibrium curve is presented in Figure 1, where U and Y denote the input and the output values of each operating point, respectively.

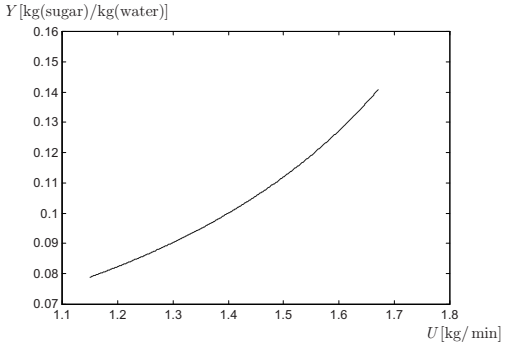


Figure 1: Operating curve of the double effect evaporator

Consider the case where identification is performed around three nominal operating points, deriving the following three corresponding linearized second order discrete-time models:

$$\textbf{Operating Point } \ell_1 = [Y_1, U_1] = [0.09395, 1.34]$$
$$S_1: \begin{aligned} &\Delta_1 y(k) - 1.7619\Delta_1 y(k-1) + 0.7756\Delta_1 y(k-2) = \\ &0.0045\Delta_1 u(k-1) - 0.0032\Delta_1 u(k-2) + \varepsilon_1(k) \end{aligned}$$
$$\textbf{Operating Point } \ell_2 = [Y_2, U_2] = [0.11066, 1.49]$$
$$S_2: \begin{aligned} &\Delta_2 y(k) - 1.7870\Delta_2 y(k-1) + 0.7978\Delta_2 y(k-2) = \\ &0.0057\Delta_2 u(k-1) - 0.0042\Delta_2 u(k-2) + \varepsilon_2(k) \end{aligned}$$
$$\textbf{Operating Point } \ell_3 = [Y_3, U_3] = [0.12902, 1.61]$$
$$S_3: \begin{aligned} &\Delta_3 y(k) - 1.8073\Delta_3 y(k-1) + 0.8160\Delta_3 y(k-2) = \\ &0.0071\Delta_3 u(k-1) - 0.0055\Delta_3 u(k-2) + \varepsilon_3(k) \end{aligned}$$

Note that the sampling period is $T = 10[\text{min}]$.

The linear models, derived through the identification procedure, are evaluated experimentally in order to determine the operating areas of the process around which the linear models constitute satisfactory approximations of the nonlinear process. The operating areas of each nominal operating point are determined by comparing the responses of the nonlinear process and the corresponding linear model for step input functions with several amplitudes. According to the above, the target and tolerance operating areas are determined according to the following [4]:

a) **Target operating areas** ($O_i, i = 1,2,3$): Consider that for $k < 0$ the process rests at an initial operating point $\rho_s = [Y_s, U_s]$ in O_i. Then, for $k \ge 0$ a suitable step input function is applied to the process such that the final operating point is $\rho_f = [Y_f, U_f]$, where ρ_f lies also within O_i. Let $V_{y,i} = [y(0) - Y_i, y(1) - Y_i, y(2) - Y_i, \cdots, y(N) - Y_i]$ be a vector with elements the deviations of the response $y(k)$ of the process from the output value Y_i of the nominal operating point ℓ_i, for the sampling instants $0,1,2,\ldots,N$, where N is selected large enough for the process to settle to its steady state value. Consider the linear model S_i, with $\varepsilon_i \equiv 0$. Apply the

step input function $\Delta_i u(k) = U_f - U_i, k \geq 0$. Using model S_i, the output signal $\Delta_i y(k), k \geq 0$ is computed considering the initial conditions: $\Delta_i y(k) = Y_s - Y_i$, $\Delta_i u(k) = U_s - U_i$ for $k < 0$. Consider the vector $V_{\Delta y,i} = [\Delta_i y(0), \Delta_i y(1), \Delta_i y(2), \cdots, \Delta_i y(N)]$. Then, for all initial and all final operating points within O_i, it should hold that $\|V_{y,i} - V_{\Delta y,i}\|_2 / \|V_{\Delta y,i}\|_2 \leq 0.15$, where $\|\cdot\|_2$ denotes the Euclidean norm.

b) **Tolerance operating areas** $\bar{O}_i, i = 1,2,3$. These areas are produced with the same procedure as O_i, except that the allowed deviation is 30%, thus $\|V_{y,i} - V_{\Delta y,i}\|_2 / \|V_{\Delta y,i}\|_2 \leq 0.30$.

The target and tolerance operating areas of the linear models $S_i, i = 1,2,3$ of the double effect evaporator are determined by the following equations:

$$O_1 = \{(u,y): U_{n,1} \leq u \leq U_{x,1}, Y_{n,1} \leq y \leq Y_{x,1}\} =$$
$$\{(u,y): 1.2200 \leq u \leq 1.4390, 0.083828 \leq y \leq 0.10435\}$$
$$\bar{O}_1 = \{(u,y): U_{l,1} \leq u \leq U_{p,1}, Y_{l,1} \leq y \leq Y_{p,1}\} =$$
$$\{(u,y): 0.8760 \leq u \leq 1.5930, 0.06404 \leq y \leq 0.12606\}$$
$$O_2 = \{(u,y): U_{n,2} \leq u \leq U_{x,2}, Y_{n,2} \leq y \leq Y_{x,2}\} =$$
$$\{(u,y): 1.3860 \leq u \leq 1.5760, 0.09852 \leq y \leq 0.12323\}$$
$$\bar{O}_2 = \{(u,y): U_{l,2} \leq u \leq U_{p,2}, Y_{l,2} \leq y \leq Y_{p,2}\} =$$
$$\{(u,y): 1.0890 \leq u \leq 1.7090, 0.07500 \leq y \leq 0.14947\}$$
$$O_3 = \{(u,y): U_{n,3} \leq u \leq U_{x,3}, Y_{n,3} \leq y \leq Y_{x,3}\} =$$
$$\{(u,y): 1.5190 \leq u \leq 1.6850, 0.11460 \leq y \leq 0.14394\}$$
$$\bar{O}_3 = \{(u,y): U_{l,3} \leq u \leq U_{p,3}, Y_{l,3} \leq y \leq Y_{p,3}\} =$$
$$\{(u,y): 1.2610 \leq u \leq 1.8010, 0.08703 \leq y \leq 0.17530\}$$

These areas are represented in Figure 2, according to which the target and tolerance operating areas of these nominal operating points satisfy the dense web principle conditions.

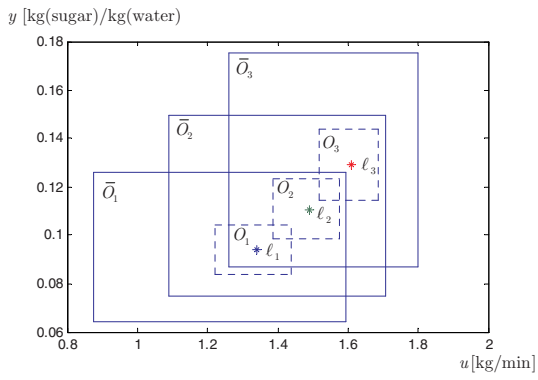


Figure 2: Target and tolerance operating areas of the three linear models

IV. Field Controller Design

A. Controller Specifications

The set of admissible controllers for each of the corresponding linearized models is determined to satisfy the specifications expressed as the following set of propositions:

$\wp_i =$ { a) The settling time of the step response of the closed loop system must be 15% better than the corresponding settling time of the linearized model S_i. b) The overshoot of the step response of the closed loop system is less than 5%. c) The steady state gain of the closed-loop system is equal to 1. d) The magnitude of the maximum eigenvalue of the closed loop system is not allowed to be more than 1% larger than the corresponding of the linearized model S_i. e) For transitions between the operating points $[Y_{n,i}, U_{n,i}]$ and $[Y_{x,i}, U_{x,i}]$, the controller output must remain within the interval $[U_{l,i}, U_{p,i}]$.}

The above specifications can be achieved by applying PI controllers in discrete-time incremental form

$$u(k) = u(k-1) + ce(k) + ge(k-1) \quad (1)$$

where $e(k) = r(k) - y(k)$ and $r(k)$ denotes the external command. Consider, for example, the case where the discrete-time description (1) is derived through backward discretization of the continuous-time PI controllers with description $u(t) = Ke(t) + (K/T_I)\int_0^t e(\tau)d\tau$. Then the parameters c, g are related with the proportional gain K, the parameter T_I and the sampling period T through the equations $c = K\left(1 + \frac{T}{T_I}\right)$, $g = -K$. Note that since $T > 0$ and $T_I > 0$, the parameters c and $-g$ have always the same sign with the gain K.

It is important to note at this point, that the design requirement (d) of $\wp_i$ allows a slight transfer of the closed-loop poles closer to the unit circle, than the open-loop poles. This is due to the fact that the PI controller may not be efficient in moving the closed-loop poles at more stable positions. The application of a SWSS algorithm for the double effect evaporator has also been studied in [8], where a more rich controller had been considered, that succeeded to improve the position of the closed-poles. However, as it is indicated by the simulation results presented in a forthcoming section, the more realistic for practical applications use of PI field controllers succeeds as well to achieve satisfactory closed-loop performance.

B. Heuristic "Common" Controller Design

Let Ω_i denote the set of controllers satisfying the design requirements $\wp_i$. The determination of the sets $\Omega_i \cap \Omega_{i+1}, i = 1,2$ of "common" controllers for the linear

models S_i and S_{i+1} can be done with the use of the heuristic "common" PI controller tuning algorithm presented in [9].

The initial area of search is determined according to the following. From the operating curve presented in Figure 1, it is obvious that the dc gain of the double effect evaporator is positive. Thus, the proportional gain K as well as the parameter c of the discrete-time description of the PI controller must be also positive, while the parameter g of the discrete-time description must be negative and with $|g|<c$ (recall the relations between c,g and the parameters K,T_I,T). Thus, in order to determine the initial search area, it suffices to determine an upper bound, let say c_m, for the parameter c. The determination of c_m is done so as to guarantee that for all transitions between any operating points of O_i and for all transitions between any operating points of O_{i+1}, the corresponding controller outputs at the first sampling instant after the controller's initialization will remain within $\bar{O}_i$ and $\bar{O}_{i+1}$, respectively.

For the case of the double effect evaporator, the initial search area for the set $\Omega_1 \cap \Omega_2$ of "common" controllers is determined by $P_{1,2}=\{(c,g)\in\mathbb{R}^2 : 0.0001\le c\le 13.0707,\ -13.0707\le g\le -0.0001\}$, while for the set $\Omega_2\cap\Omega_3$ of "common" controllers is determined by $P_{2,3}=\{(c,g)\in\mathbb{R}^2 : 0.0001\le c\le 9.6122,\ -9.6122\le g\le -0.0001\}$.

The set of "common" controllers $\Omega_i\cap\Omega_{i+1}$ is determined under an additional requirement, besides those determined by $\wp_i$ and $\wp_{i+1}$. More specifically, we additionally require that for all transitions between the operating points $[Y_{n,i},U_{n,i}]$ and $[Y_{x,i+1},U_{x,i+1}]$ the controller output remains within the interval $[U_{l,i},U_{p,i+1}]$. This additional requirement implies that during transitions between operating points lying in $O_i\cup O_{i+1}$, the plant's I/O trajectories remain within $\bar{O}_i\cup\bar{O}_{i+1}$.

Among the controllers of $\Omega_1\cap\Omega_2$, we choose one that minimizes the quantity $\max\{st_1,st_2\}$, where $st_i, i=1,2$ is the settling time of the step response of the closed loop system produced by the application of the controller to the linear models S_1 and S_2, respectively. According to this criterion, from the set $\Omega_1\cap\Omega_2$, we select to use the controller that corresponds to $(c,g)=(8.6650,-7.8421)$. In a similar way, from the set $\Omega_2\cap\Omega_3$ we select to use the controller that corresponds to $(c,g)=(6.9208,-6.3440)$.

Note that the application of "common" controllers for adjacent operating points is a robust control design approach, that may be used to extend the operating range of the plant, where the field controller is effective. However, this approach becomes inadequate, in general, as the extend of the operating area covered by the nominal operating points increases, since in this case it may not be possible to achieve the design requirements simultaneously for all operating points under consideration. In these cases "common" controller design has to be combined with controller switching, so as to cover the whole operating range.

Indeed, the design of "common" PI controllers for the double effect evaporator has also been studied in [9], where a "common" PI controller has been designed for three operating points, so as to achieve the design requirements (a)-(d) of $\wp_i$. However, even an extensive search failed to achieve the design (with the same design requirements) of a "common" PI controller, for all three operating points ℓ_1, ℓ_2 and ℓ_3, determined in Section II, which cover a wider operating range, than the corresponding of [9]. This inability is treated by the application of the supervisory switching controller that performs switching between PI controllers "common" for pairs of adjacent operating points.

V. Simulation Results

A. Step-Wise Safe Switching Control

Consider now that the nonlinear description of the double effect evaporator is unknown to the designer, while the only available information are the linearizations $S_i, i=1,2,3$ around the corresponding nominal operating points, as well as the operating areas $O_i,\bar{O}_i$, $i=1,2,3$. Then, a safe switching controller can be applied in order to control the double effect evaporator process.

In order to move the system from the nominal operating point ℓ_1 to the nominal operating point ℓ_3, the designer applies the safe switching step algorithm that takes the following form:

a) Initially the designer selects a controller from the common set $\Omega_1\cap\Omega_2$ of regional stabilizing controllers for the nominal operating points ℓ_1 and ℓ_2. Specifically, the applied PI controller is:

$$C_{1,2}:\quad u(k+1)=u(k)+8.6650e(k+1)-7.8421e(k)$$

with $r(k)=Y_{0,2}$. The selected external command drives the system from the nominal operating point ℓ_1 to the operating point ℓ_2.

b) As soon as the system trajectory approaches the nominal operating point ℓ_2, the designer switches controller. The new controller is selected from the common set $\Omega_2\cap\Omega_3$ of regional stabilizing controllers for the nominal operating points ℓ_2 and ℓ_3. Specifically, the applied PI controller is:

$$C_{2,3}: u(k+1)=u(k)+6.9208e(k+1)-6.3440e(k)$$

with $r(k)=Y_{0,3}$. The selected external command drives the system from the nominal operating point ℓ_2 to the nominal operating point ℓ_3.

Note that controller switching takes place provided that the following two conditions are satisfied for three sequential instants of time

$$\left|y(k)-Y_2\right| \le 0.05\left|Y_2-Y_1\right|, \left|u(k)-U_2\right| \le 0.05\left|U_2-U_1\right| \quad (2)$$

Figure 3 presents the simulation results derived from the application of the aforementioned switching strategy, when the double effect evaporator process is described by the following nonlinear state equation ([10], [11])

$$\begin{aligned} \dot{x}_1(t) &= d_1F_0(C_0-x_1(t))+d_2x_1(t)u(t) \\ \dot{x}_2(t) &= d_3F_0(x_1(t)-x_2(t))+(d_4x_1(t)+d_5x_2(t))u(t) \end{aligned} \quad (3)$$

where $y(t)=x_2(t)$, x_1 and x_2 are the output concentrations of the first and the second effect respectively, $F_0=2.525$[kg/min] is the feed flow to the first effect, $C_0=0.04$[kg(sugar)/kg(water)] is the feed concentration to the first effect and $d_1,\ldots,d_5$ are process parameters with values $d_1=0.010526$[1/kg], $d_2=0.008510$[1/kg], $d_3=0.009524$[1/kg], $d_4=-0.007700$[1/kg], $d_5=0.010306$[1/kg]. Note that the controller output is applied to the process with the use of a zero-order hold.

In the same figure, we also present, for comparison reasons, the corresponding response of the nonlinear model for input function determined by

$$u(t)=\begin{cases} U_2 & ,t<k_0T \\ U_3 & ,t\ge k_0T \end{cases} \quad (4)$$

Switching takes place at k_0T provided that the output $y(t)$ of the process satisfies the first of the inequalities (2) for $k=k_0-2,k_0-1,k_0$. It is obvious from Figure 3 that the application of the switching controller results in a very good improvement of settling time, with a small increase of the overshoot.

Figure 4 presents the corresponding controller output. It can be noticed from the curve of Figure 4, that the improvement of the system response is accomplished with an increase of the controller's output values in comparison with those determined by (4), fact that was well expected. However, the restrictions imposed on the selection of the PI controllers guarantee that the (u,y) trajectories at each step of the algorithm remain within the corresponding tolerance operating areas at the transient stage and within the corresponding target operating areas at the steady state stage. This is also made clear by the representation of the (u,y) trajectories of Figure 5.

B. Comparison with Other Switching Controllers

The advantages of the safe switching algorithm in comparison with a specific standard switching strategy is studied in the current section. Figure 6 presents a state transition from the operating point $[Y_{n,1},U_{n,1}]$ to the operating point $[Y_{x,2},U_{x,2}]$ using two different approaches:

a) Application of a PI controller belonging to the set $\Omega_1\cap\Omega_2$ (safe switching algorithm)

b) Switching between a PI controller of the set Ω_1 and a PI controller of the set Ω_2. The switching takes place when the output of the double effect evaporator satisfies the condition $y(k)>(Y_{n,2}+Y_{x,1})/2$. Note that the determination of the sets Ω_1 and Ω_2 is made by searching within the areas ([9])

$$\begin{aligned} P_1 = \{(c,g)\in\ ^2 &: 0.0001\le c\le 18.1732, \\ &-18.1732\le g\le -0.0001\}, \\ P_2 = \{(c,g)\in\ ^2 &: 0.0001\le c\le 13.0707, \\ &-13.0707\le g\le -0.0001\} \end{aligned}$$

respectively, which are determined by selecting upper bounds for the parameter c, so as to guarantee that for all transitions between any operating points of O_i, O_{i+1}, respectively, the corresponding controller outputs at the first sampling instant after the controller's initialization will remain within $\bar{O}_i, \bar{O}_{i+1}$ respectively. The final selection of the controllers are made using the same criteria with those used for safe switching, with the only difference that now the criteria are applied for each linear model separately. Thus, for the linear model S_1 we select the controller with parameters $(c,g)=(17.6280,-15.9924)$, while for the linear model S_2 we use the controller with parameters $(c,g)=(12.2865,-11.2407)$.

It is obvious from Figure 6, that application of a "common" controller results in much better performance than switching (without being close to steady state) between two controllers each of which is suitable for only one of the two operating points.

VI. Conclusions

The Step-Wise Safe Switching approach for SISO systems has been applied to a second-order nonlinear process. The performance of SWSS has been studied using closed-loop simulation results for transitions of the I/O trajectories between operating areas of three adjacent operating points, satisfying the dense web principle. According to the presented simulation results, the SWSS algorithm achieves satisfactory and safe closed-loop performance. Moreover, the performance of the SWSS scheme seems to be advantageous when compared to a standard switching algorithm, as well as when compared to an approach seeking for "common" controllers for more than two operating points.

Acknowledgment

The present work is co-financed by the Hellenic Ministry of Education and Religious Affairs' and the ESF of the European Union within the framework of the "Operational Programme

for Education and Initial Vocational Training" (Operation "Archimedes-I").

REFERENCES

[1] A. Leonessa, W.M. Haddad, V. Chelaboina, "Nonlinear system stabilization via hierarchical switching control", *IEEE Trans. on Autom. Control*, vol. 46, pp. 17-28, 2001

[2] M.W. McConley, B.D. Appleby, M.A. Dalheh, E. Feron, "A computationally efficient Lyapunov-based scheduling procedure for control of nonlinear systems with stability guarantees", *IEEE Trans. on Autom. Control*, vol. 45, pp. 33-49, 2000

[3] E.F. Costa, V.A. Oliveira, "Gain scheduled controllers for dynamic systems using sector nonlinearities", *Automatica*, vol.38, pp. 1247-1250, 2002

[4] F.N. Koumboulis, R.E. King, A. Stathaki, "Logic-Based Switching Controllers – A stepwise safe switching approach", *Information Sciences*, vol. 177, pp. 2736–2755, 2007

[5] F.N. Koumboulis, M.P. Tzamtzi, "Towards Analytic Solutions of Step-Wise Safe Switching for Known Affine-Linear Models", *Int. Conf. on Num. Anal. App. Math. 2007 (ICNAAM 2007)*, Corfu, Greece, Sep. 2007

[6] F.N. Koumboulis, M.P. Tzamtzi, "On the Stability of the Step-Wise Safe Switching Control Approach", submitted

[7] F.N. Koumboulis, M.P. Tzamtzi, "Supervisory Scheme for Stepwise Safe Switching Controllers", *WASET Intern. Conf. on Control, Autom. and Systems (ICCAS 2007)*, Venice, Italy, 2007, pp. 244-251.

[8] F.N. Koumboulis, M.P. Tzamtzi, "Multivariable Step-Wise Safe Switching Controllers", *Int. Conf. on Comp. Intelligence for Modeling, Control and Automation 2005*, Austria, vol. II, pp. 624-630, 2005.

[9] F. N. Koumboulis, "On the heuristic design of common PI controllers for multi-model plants", *10th IEEE International Conf. on Emerging Techn. and Factory Automation (ETFA 2005)*, Italy, pp. 975-982, 2005

[10] H. Sira-Ramirez, O.Llanes-Santiago, "Dynamical Discontinuous Feedback Strategies in the Regulation of Nonlinear Chemical Processes", *IEEE Trans. Control Systems Techn.*, vol. 2, pp. 11-21, 1994

[11] A. Montano, G. Silva, "Design of a nonlinear control for a double effect evaporator", pp. 2256-2261, *ECC 1991*, Grenoble, France,July 2-5, 1991.

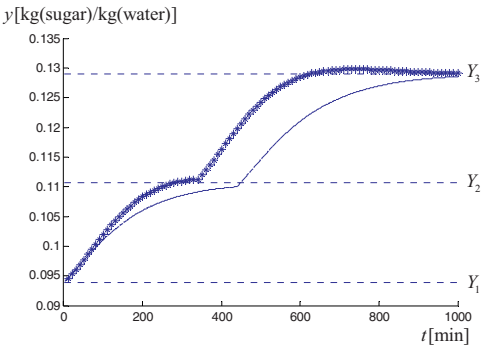


Figure 3: Open loop (-) and closed loop trajectories (*)

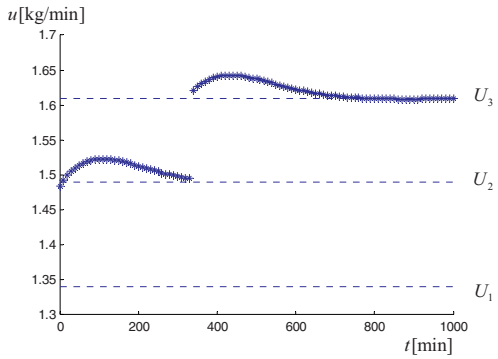


Figure 4: Switching controller output

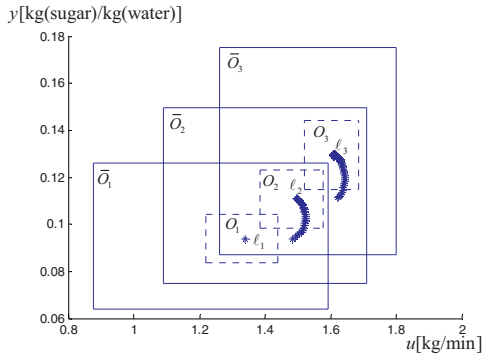


Figure 5: Representation of closed loop trajectories in the (u, y) plane

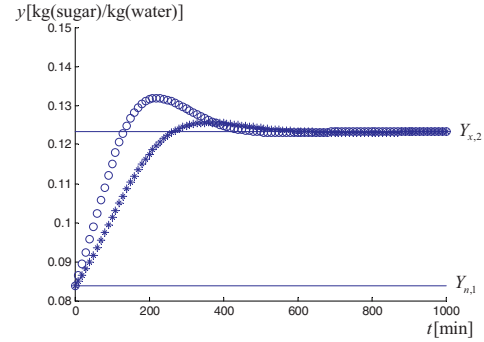


Figure 6: Comparison of safe switching with a standard switching algorithm ('*': Safe switching, 'o': Standard switching algorithm)

Use of a Connection Model for Dynamic Systems

M. Braae
Department of Electrical Engineering
University of Cape Town
Rondebosch, 7701 South Africa.

Abstract - **An algebraic model for interconnected dynamic systems with dead time is proposed. The model structure separates the system dynamics and connections into two sets of equations in which the dynamic equation is invariant under changes in system interconnections. Useful properties of the characteristic connection matrix of the resulting model are illustrated through applications of the theoretical results for feedback connections and for model inversion. It is shown that the former can guide the design of control structures in industrial plants before modelling their dynamics whilst the latter can optimize block diagram connections to ensure causal dynamic blocks where possible.**

I. INTRODUCTION

Industrial plants with distributed, dynamic sub-systems can be approximated in a range of theoretical frameworks for mathematical analysis [1]. The structures and dynamics of such plants and their associated control systems are often combined to form a single dynamic model for the interconnected system. The dynamics of this model are then analyzed.

The state space formulation:

$$\begin{aligned} \frac{d}{dt}\mathbf{x}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \end{aligned} \tag{1}$$

where $\mathbf{A} \in \mathbf{R}^{n\times n}$, $\mathbf{B} \in \mathbf{R}^{n\times m}$, $\mathbf{C} \in \mathbf{R}^{l\times n}$ and $\mathbf{D} \in \mathbf{R}^{l\times m}$ is a popular theoretical framework even though the inclusion of dead-time terms leads to increased complexity, especially in distributed control systems [2].

The transfer function matrix model:

$$\mathbf{y}(s) = \mathbf{G}(s)\mathbf{u}(s) \tag{2}$$

is another common formulation [3]. This readily incorporates dead-time terms in its matrix elements, though its frequency response version may be preferred for industrial plants where material handling introduces significant dead-times in every pipe and on every conveyor belt.

If interconnections between plant sub-systems become the focus of an analysis then the state space model can be re-written in a form that emphasizes such connections. Specifically the dynamics of each sub-system can be represented by equations such as:

$$\mathbf{S:} \quad \frac{d}{dt}\mathbf{x}_i(t) = \mathbf{A}_i\mathbf{x}_i + \mathbf{B}_i\mathbf{u}_i + \sum_{j=1}^{N} \mathbf{e}_{ij}\mathbf{A}_{ij}\mathbf{x}_j, \quad i \in N \tag{3}$$

or

$$\mathbf{S:} \quad \frac{d}{dt}\mathbf{x}_i(t) = \mathbf{A}_i\mathbf{x}_i + \mathbf{B}_i(\mathbf{u}_i + \sum_{j=1}^{N} \mathbf{e}_{ij}\mathbf{D}_{ij}\mathbf{x}_j), \quad i \in N \tag{4}$$

where the connection matrix $\mathbf{E}$ contains elements that are either binary, $e_{ij} \in \{0,1\}$, or fractional, $e_{ij} \in [0,1]$ [4]. Note the implied element-by-element multiplication for matrix $\mathbf{E}$ in both (3) and (4).

Variations of the distributed system formulation abound e.g. [2] and [4] to [11] inclusive. Some formulations use transfer function models but most opt for state space formats, even when dealing with input-output structures [12].

II. PROPOSED CONNECTION MODEL

A generalization of the state space model of (1) yields a set of equations in which the dynamic components and the connection components of a dynamic system are separated into two distinct sets of equations [13] – One set defines the system dynamics; the other its interconnections.

The resulting Transfer Function Connection Matrix (TFCM) model has the form:

$$\begin{aligned} \mathbf{z} &= \mathbf{M}(s)\mathbf{v} \\ \mathbf{v} &= \mathbf{X}\mathbf{z} + \mathbf{U}\mathbf{u} \\ \mathbf{y} &= \mathbf{Y}\mathbf{z} + \mathbf{W}\mathbf{u} \end{aligned} \tag{5}$$

where $\mathbf{M}(s)$ is a transfer function matrix [3] that is generally formulated to be diagonal or block diagonal in the case of interconnected systems. Unlike state space systems it can readily contain dead-time terms. The four connections matrices, $\mathbf{X}$, $\mathbf{U}$, $\mathbf{Y}$ and $\mathbf{W}$, contain constant elements that define the connections between signals in the system. Of these the characteristic connection matrix, $\mathbf{X}$, is of most interest as it determines useful properties of the interconnected system without reference to its dynamics [13].

The inputs to each dynamic sub-system are contained in the vector $\mathbf{v}$ whilst the outputs are in vector $\mathbf{z}$. The *manipulated* (or *actuated* or *controlled* or *input*) variables of the entire system are in vector $\mathbf{u}$ and the *measured* (or *sensed* or *output*) variables are in vector $\mathbf{y}$.

In its elemental form matrix $\mathbf{M}(s)$ is a diagonal matrix of SISO transfer functions. Elimination of the internal variables, $\mathbf{v}$ and $\mathbf{z}$, reveals that (5) contains a closed loop system relating its sensor variables, $\mathbf{y}$, to its

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 168–173.

actuator variables, $\mathbf{u}$. (This is similar to the traditional state space model of (1) from which it was derived. In both cases the positive feedback convention is used.)

III. STABILITY

Transfer function theory for MIMO closed loop systems [3] shows that the stability of the TFCM model is given by:

$$\phi_c(s) = \pm\left|\mathbf{I} - \mathbf{M}(s)\mathbf{X}\right|\phi_o(s) \tag{6}$$

where $\phi_c(s)$ is the characteristic function of the TFCM model and $\phi_o(s)$ is the product of all the characteristic functions of its sub-systems. The latter are readily obtained from the dynamic models in $\mathbf{M}(s)$, especially in the elemental case.

IV. DISTRIBUTED STABILITY

For distributed systems the dynamic model, $\mathbf{M}(s)$, will be block diagonal, similar to the 2-block form:

$$\begin{aligned}
\begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\end{bmatrix} &= \begin{bmatrix}\mathbf{M}_{11}(s) & \mathbf{0}\\ \mathbf{0} & \mathbf{M}_{22}(s)\end{bmatrix}\begin{bmatrix}\mathbf{v}_1\\ \mathbf{v}_2\end{bmatrix}\\
\begin{bmatrix}\mathbf{v}_1\\ \mathbf{v}_2\end{bmatrix} &= \begin{bmatrix}\mathbf{X}_{11} & \mathbf{X}_{12}\\ \mathbf{X}_{21} & \mathbf{X}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\end{bmatrix} + \begin{bmatrix}\mathbf{U}_1\\ \mathbf{U}_2\end{bmatrix}\mathbf{u}\\
\mathbf{y} &= \begin{bmatrix}\mathbf{Y}_1 & \mathbf{Y}_2\end{bmatrix}\begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\end{bmatrix} + \mathbf{W}\mathbf{u}
\end{aligned} \tag{7}$$

The stability of this conformally partitioned TFCM system is determined from its characteristic function:

$$\phi_c(s) = \pm\left|\begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix} - \begin{bmatrix}\mathbf{M}_{11}(s) & \mathbf{0}\\ \mathbf{0} & \mathbf{M}_{22}(s)\end{bmatrix}\begin{bmatrix}\mathbf{X}_{11} & \mathbf{X}_{12}\\ \mathbf{X}_{21} & \mathbf{X}_{22}\end{bmatrix}\right|\phi_o(s) \tag{8}$$

or, after performing the matrix multiplication:

$$\phi_c(s) = \pm\left|\begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix} - \begin{bmatrix}\mathbf{M}_{11}(s)\mathbf{X}_{11} & \mathbf{M}_{11}(s)\mathbf{X}_{12}\\ \mathbf{M}_{22}(s)\mathbf{X}_{21} & \mathbf{M}_{22}(s)\mathbf{X}_{22}\end{bmatrix}\right|\phi_o(s) \tag{9}$$

If the characteristic connection matrix, $\mathbf{X}$, is lower block triangular then (9) reduces to:

$$\phi_c(s) = \pm\left|\begin{bmatrix}\mathbf{I} - \mathbf{M}_{11}(s)\mathbf{X}_{11} & \mathbf{0}\\ -\mathbf{M}_{22}(s)\mathbf{X}_{21} & \mathbf{I}_{z2} - \mathbf{M}_{22}(s)\mathbf{X}_{22}\end{bmatrix}\right|\phi_o(s) \tag{10}$$

The determinant of this partitioned matrix can be expressed as [3]:

$$\phi_c(s) = \pm\left|\mathbf{I} - \mathbf{M}_{11}(s)\mathbf{X}_{11}\right|\left|\mathbf{I} - \mathbf{M}_{22}(s)\mathbf{X}_{22}\right|\phi_o(s) \tag{11}$$

From this it can be seen that the stability of the TFCM model for a distributed control system with a lower triangular characteristic connection matrix can be deduced from the characteristic functions of its two sub-systems:

$$\phi_c(s) = \pm\frac{\phi_{c1}(s)}{\phi_{o1}(s)} \times \frac{\phi_{c2}(s)}{\phi_{o2}(s)}\phi_o(s) = \pm\phi_{c1}(s)\phi_{c2}(s) \tag{12}$$

An identical result is readily deduced for a TFCM model with an upper triangular characteristic connection matrix.

If in addition to being lower (or upper) block triangular, one of the blocks on the diagonal of the characteristic connection matrix is zero then the corresponding dynamic models will remain in open loop within the interconnected system.

For example if $\mathbf{X}_{22} = \mathbf{0}$ then (11) reduces to:

$$\phi_c(s) = \pm\left|\mathbf{I} - \mathbf{M}_{11}(s)\mathbf{X}_{11}\right|\phi_o(s) = \pm\phi_{c1}(s)\phi_{o2}(s) \tag{13}$$

and clearly illustrates the effect.

The above results reveal the significance of the characteristic connection matrix, $\mathbf{X}$, in the TFCM model in that its structure impacts directly on the stability analysis required for the given plant.

V. FEEDBACK CONNECTIONS

The measured and manipulated signals in the system defined by (5) can be connected in a control loop that is described by the feedback connection equation:

$$\mathbf{u} = \mathbf{P}\mathbf{r} - \mathbf{F}\mathbf{y} \tag{14}$$

where vector $\mathbf{r}$ contains the system setpoint variables while matrices $\mathbf{P}$ and $\mathbf{F}$ are the connections matrices to the inputs $\mathbf{u}$ from the system setpoints $\mathbf{r}$ and outputs $\mathbf{y}$ respectively.

Substitution of (14) into (5) followed by further manipulation yields the TFCM for the controlled system:

$$\begin{aligned}
\mathbf{z} &= \mathbf{M}(s)\mathbf{v}\\
\mathbf{v} &= -\{\mathbf{X} + \mathbf{U}\mathbf{F}\left[\mathbf{I} + \mathbf{W}\mathbf{F}\right]^{-1}\mathbf{Y}\}\mathbf{z}\\
&\quad + \{\mathbf{U}\mathbf{P} + \mathbf{U}\mathbf{F}\left[\mathbf{I} + \mathbf{W}\mathbf{F}\right]^{-1}\mathbf{W}\mathbf{P}\}\mathbf{r}\\
\mathbf{y} &= \{\left[\mathbf{I} + \mathbf{W}\mathbf{F}\right]^{-1}\mathbf{Y}\}\mathbf{z}\\
&\quad + \{\left[\mathbf{I} + \mathbf{W}\mathbf{F}\right]^{-1}\mathbf{W}\mathbf{P}\}\mathbf{r}
\end{aligned} \tag{15}$$

It should be noted that the connection matrices only are altered by the feedback connection and that the dynamic matrix equation for the sub-systems has not been affected. This invariance of the dynamics to the interconnection of systems is a key feature of the TFCM models. It is also apparent from the matrix inversions in the final system that the elements of the connection matrices will be real rather than binary or fractional constants as in (3) and (4). The matrix inversion in (15) needs to be possible and may necessitate a change in the feedback connections, $\mathbf{F}$.

VI. INVERSE TFCM MODEL

The inverse TFCM is readily deduced from (5) to be:

$$\begin{aligned}
\mathbf{v} &= \hat{\mathbf{M}}(s)\mathbf{z}\\
\mathbf{z} &= \{\hat{\mathbf{X}}\}\mathbf{v} - \{\hat{\mathbf{X}}\mathbf{U}\}\mathbf{u}\\
\mathbf{y} &= \{\mathbf{Y}\hat{\mathbf{X}}\}\mathbf{v} + \{\mathbf{W} - \mathbf{Y}\hat{\mathbf{X}}\mathbf{U}\}\mathbf{u}
\end{aligned} \tag{16}$$

provided that both the dynamic model and the characteristic connection matrix are non-singular. (The notation uses symbol $\hat{\mathbf{M}}$ as the inverse of matrix $\mathbf{M}$.)

In practice the dynamic model is (block) diagonal and hence invertible though causal dynamic models may become non-causal, and vice-versa. However the characteristic connection matrix is often singular. Thus

(16) needs to be modified to allow partial inversion of some components of a partitioned TFCM model. The characteristic connection matrix of one of the diagonal blocks should contain a square matrix that is non-singular.

Assume a partition of a TFCM in (5) given by:

$$\begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\end{bmatrix}=\begin{bmatrix}\mathbf{M}_{11}(s) & \mathbf{M}_{12}(s)\\ \mathbf{M}_{21}(s) & \mathbf{M}_{22}(s)\end{bmatrix}\begin{bmatrix}\mathbf{v}_1\\ \mathbf{v}_2\end{bmatrix}$$
$$\begin{bmatrix}\mathbf{v}_1\\ \mathbf{v}_2\end{bmatrix}=\begin{bmatrix}\mathbf{X}_{11} & \mathbf{X}_{12}\\ \mathbf{X}_{21} & \mathbf{X}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\end{bmatrix}+\begin{bmatrix}\mathbf{U}_1\\ \mathbf{U}_2\end{bmatrix}\mathbf{u} \qquad (17)$$
$$\mathbf{y}=\begin{bmatrix}\mathbf{Y}_1 & \mathbf{Y}_2\end{bmatrix}\begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\end{bmatrix}+\mathbf{W}\mathbf{u}$$

in which the characteristic connection matrix $\mathbf{X}_{11}(s)$ and the dynamic model matrix $\mathbf{M}_{11}(s)$ are invertible. For elemental dynamic models the latter sub-matrix will be diagonal and so the individual transfer functions are likely to be a mix of causal and non-causal SISO systems.

The partially inverted TFCM model that results is:

$$\begin{bmatrix}\mathbf{v}_1\\ \mathbf{z}_2\end{bmatrix}=\begin{bmatrix}\hat{\mathbf{M}}_{11}(s) & -\hat{\mathbf{M}}_{11}(s)\mathbf{M}_{12}(s)\\ \mathbf{M}_{21}(s)\hat{\mathbf{M}}_{11}(s) & \mathbf{M}_{22}(s)-\mathbf{M}_{21}(s)\hat{\mathbf{M}}_{11}(s)\mathbf{M}_{12}(s)\end{bmatrix}\begin{bmatrix}\mathbf{z}_1\\ \mathbf{v}_2\end{bmatrix}$$

$$\begin{bmatrix}\mathbf{z}_1\\ \mathbf{v}_2\end{bmatrix}=\begin{bmatrix}\hat{\mathbf{X}}_{11} & -\hat{\mathbf{X}}_{11}\mathbf{X}_{12}\\ \mathbf{X}_{21}\hat{\mathbf{X}}_{11} & \mathbf{X}_{22}-\mathbf{X}_{21}\hat{\mathbf{X}}_{11}\mathbf{X}_{12}\end{bmatrix}\begin{bmatrix}\mathbf{v}_1\\ \mathbf{z}_2\end{bmatrix}+\begin{bmatrix}-\hat{\mathbf{X}}_{11}\mathbf{U}_1\\ \mathbf{U}_2-\mathbf{X}_{21}\hat{\mathbf{X}}_{11}\mathbf{U}_1\end{bmatrix}\mathbf{u}$$

$$\mathbf{y}=\begin{bmatrix}\mathbf{Y}_1\hat{\mathbf{X}}_{11} & \mathbf{Y}_2-\mathbf{Y}_1\hat{\mathbf{X}}_{11}\mathbf{X}_{12}\end{bmatrix}\begin{bmatrix}\mathbf{v}_1\\ \mathbf{z}_2\end{bmatrix}+\{\mathbf{W}-\mathbf{Y}_1\hat{\mathbf{X}}_{11}\mathbf{U}_1\}\mathbf{u} \qquad (18)$$

For systems with a block diagonal dynamic model this reduces to:

$$\begin{bmatrix}\mathbf{v}_1\\ \mathbf{z}_2\end{bmatrix}=\begin{bmatrix}\hat{\mathbf{M}}_{11}(s) & \mathbf{0}\\ \mathbf{0} & \mathbf{M}_{22}(s)\end{bmatrix}\begin{bmatrix}\mathbf{z}_1\\ \mathbf{v}_2\end{bmatrix}$$

$$\begin{bmatrix}\mathbf{z}_1\\ \mathbf{v}_2\end{bmatrix}=\begin{bmatrix}\hat{\mathbf{X}}_{11} & -\hat{\mathbf{X}}_{11}\mathbf{X}_{12}\\ \mathbf{X}_{21}\hat{\mathbf{X}}_{11} & \mathbf{X}_{22}-\mathbf{X}_{21}\hat{\mathbf{X}}_{11}\mathbf{X}_{12}\end{bmatrix}\begin{bmatrix}\mathbf{v}_1\\ \mathbf{z}_2\end{bmatrix}+\begin{bmatrix}-\hat{\mathbf{X}}_{11}\mathbf{U}_1\\ -\mathbf{X}_{21}\hat{\mathbf{X}}_{11}\mathbf{U}_1+\mathbf{U}_2\end{bmatrix}\mathbf{u} \qquad (19)$$

$$\mathbf{y}=\begin{bmatrix}\mathbf{Y}_1\hat{\mathbf{X}}_{11} & \mathbf{Y}_2-\mathbf{Y}_1\hat{\mathbf{X}}_{11}\mathbf{X}_{12}\end{bmatrix}\begin{bmatrix}\mathbf{v}_1\\ \mathbf{z}_2\end{bmatrix}+\{\mathbf{W}-\mathbf{Y}_1\hat{\mathbf{X}}_{11}\mathbf{U}_1\}\mathbf{u}$$

In practice it would always be possible to formulate the original system in terms of (19) rather than (18).

VII. APPLICATIONS

The use of TFCM models is illustrated by two examples. Both show the importance of the characteristic connection matrix.

In the first application this matrix guides loop pairing in an industrial plant without requiring its dynamic models. In the second application the characteristic connection matrix identifies non-causal units within a block diagram that can be converted to causal units by partial inversion.

A. An Interconnected Plant Model

Consider an industrial-scale plant containing twenty-seven dynamic sub-systems that form a diagonal dynamic matrix, $\mathbf{M}(s)$, of SISO transfer functions.

The TFCM model for the plant is given in (20) in which one connection matrix is shown as a transposed matrix.

$$\mathbf{z}=\mathbf{M}(s)\mathbf{v}$$
$$\mathbf{v}=\mathbf{0}\mathbf{z}+\begin{bmatrix}1&0&0&0&0&0&0\\1&0&0&0&0&0&0\\0&1&0&0&0&0&0\\0&1&0&0&0&0&0\\0&1&0&0&0&0&0\\0&1&0&0&0&0&0\\0&1&0&0&0&0&0\\0&0&1&0&0&0&0\\0&0&1&0&0&0&0\\0&0&1&0&0&0&0\\0&0&0&1&0&0&0\\0&0&0&1&0&0&0\\0&0&0&1&0&0&0\\0&0&0&1&0&0&0\\0&0&0&1&0&0&0\\0&0&0&0&1&0&0\\0&0&0&0&1&0&0\\0&0&0&0&1&0&0\\0&0&0&0&1&0&0\\0&0&0&0&0&1&0\\0&0&0&0&0&1&0\\0&0&0&0&0&1&0\\0&0&0&0&0&1&0\\0&0&0&0&0&1&0\\0&0&0&0&0&0&1\\0&0&0&0&0&0&1\end{bmatrix}\mathbf{u} \qquad (20)$$

$$\mathbf{y}=\begin{bmatrix}1&0&0&0&0&0&0\\0&0&1&0&0&0&0\\1&0&0&0&0&0&0\\0&1&0&0&0&0&0\\0&0&1&0&0&0&0\\0&0&0&1&0&0&0\\0&0&0&0&0&0&1\\1&0&0&0&0&0&0\\0&0&1&0&0&0&0\\0&0&0&0&0&0&1\\1&0&0&0&0&0&0\\0&1&0&0&0&0&0\\0&0&0&1&0&0&0\\0&0&0&0&1&0&0\\0&0&0&0&0&0&1\\1&0&0&0&0&0&0\\0&0&1&0&0&0&0\\0&0&0&1&0&0&0\\0&0&0&0&1&0&0\\0&1&0&0&0&0&0\\0&0&1&0&0&0&0\\0&0&0&1&0&0&0\\0&0&0&0&1&0&0\\0&0&0&0&0&1&0\\0&0&0&0&0&0&1\\1&0&0&0&0&0&0\\0&0&1&0&0&0&0\end{bmatrix}^{\mathrm{T}}\mathbf{z}+\mathbf{W}\mathbf{u}$$

Connecting the seven outputs back to corresponding inputs by (14) requires the feedback connection matrix:

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{21}$$

The characteristic connection matrix for the resulting TFCM is derived from (15) and given in Fig.1 (in which the first row and column show the permutations of the model orderings).

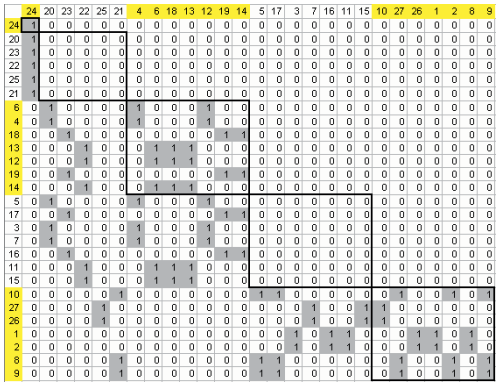


Fig.1. Characteristic Connection Matrix for Initial Feedback Connections

By the diagonal blockings outlined as boxes in Fig.1, this matrix indicates that model {24} will form a SISO closed loop system under the feedback proposed in (21), whilst models {4, 6, 18, 13, 12, 19, 14} and {10, 27, 26, 1, 2, 8, 9} will form two MIMO closed loop systems. The other models {20, 23, 22, 25, 21} and {5, 17, 3, 7, 16, 11, 15} remain in open loop since the associated block diagonals in the characteristic connection matrix are zero.

Changing a single loop pairing that alters the original feedback connection matrix of (21) to:

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{22}$$

yields a TFCM that has a characteristic connection matrix shown in Fig. 2. (Once again the model permutations are given in the first column and first row.)

This matrix shows that models {24} and {10} will form SISO closed loop systems, whilst models {4, 6, 12, 13, 14, 18, 19} and models {26, 27, 1, 2} form two MIMO closed loop systems. The other models {20, 22, 23, 21, 25}, {7, 3, 5, 15, 11, 16, 17} and {8, 9} remain in open loop. Thus the number of simple SISO loops has increased to two, one MIMO system is smaller and only 13 of the 27 dynamic models need to be derived.

In industrial plants, where the cost of modelling dynamics is high [14], the characteristic connection matrix is readily deduced, even for very complex plants. Therefore this result provides a useful engineering technique for designing the structure of its control system (especially for finding SISO loops) and for guiding the system identification phase of large-scale process control projects.

A direct comparison of Fig. 1 and Fig. 2 also shows that the single change in (21) to form (22) has resulted in a characteristic connections matrix that now has many more connections between its sub-systems (as indicated by the shaded cells).

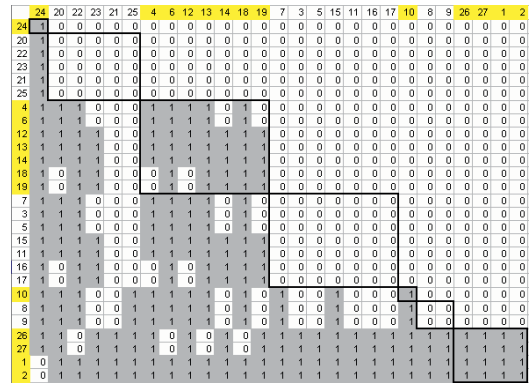


Fig. 2. Characteristic Connection Matrix for Improved Feedback Connections

B. Inverse Simulation Model

With reference to (19), the inversion of a partitioned TFCM model would allow non-causal models of sub-systems in an original dynamic model, $\mathbf{M}(s)$, to be inverted to form causal models in all the dynamic blocks of the inverse TFCM model.

To illustrate the use of this result, consider the simple R-L-C electric circuit shown in Fig.3.

A block diagram for this circuit is given in Fig.4 (where shaded segments of comparators indicate signal negation). The second block, marked #2, clearly contains a non-causal transfer function that could benefit from inversion to form a causal block for digital simulation studies.

The algebraic equations for the TFCM model of the RLC circuit are readily produced by inspection of Fig.4.

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{R} & 0 & 0 \\ 0 & sL & 0 \\ 0 & 0 & \frac{1}{sC} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u \tag{23}$$

$$y = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} + [\,0\,]\,u$$

The (2,2) element of the characteristic connection matrix in (23) corresponds to the inductance that is represented by a non-causal model, sL, in Fig.4. Unfortunately this (2,2) element is zero and hence not an invertible 1x1 sub-matrix of $\mathbf{X}$. Thus the single non-

causal block in Fig.4 cannot be converted to a causal block by inversion of such a simple partitioning of the TFCM model.

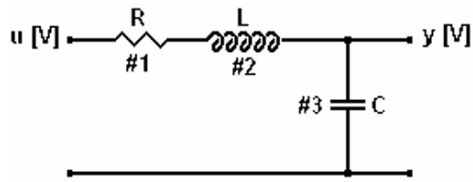


Fig. 3. Schematic of Simple R-L-C Circuit

However by partitioning the characteristic connection matrix differently:

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \left[\begin{array}{cc:c} \frac{1}{R} & 0 & 0 \\ 0 & sL & 0 \\ \hdashline 0 & 0 & \frac{1}{sC} \end{array}\right] \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \left[\begin{array}{cc:c} 0 & -1 & -1 \\ 1 & 0 & 0 \\ \hdashline 1 & 0 & 0 \end{array}\right] \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u \qquad (24)$$

$$y = \left[\begin{array}{cc:c} 0 & 0 & 1 \end{array}\right] \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} + [0]u$$

the (1,1) block diagonal sub-matrix in the characteristic connection matrix forms a 2x2 non-singular matrix that can be inverted. Hence the inverse TFCM model exists for this case and can be deduced from (19) to be:

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \left[\begin{array}{cc:c} R & 0 & 0 \\ 0 & \frac{1}{sL} & 0 \\ \hdashline 0 & 0 & \frac{1}{sC} \end{array}\right] \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \left[\begin{array}{cc:c} 0 & 1 & 0 \\ -1 & 0 & -1 \\ \hdashline 0 & 1 & 0 \end{array}\right] \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} u \qquad (25)$$

$$y = \left[\begin{array}{cc:c} 0 & 0 & 1 \end{array}\right] \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} + [0]u$$

in which the dynamics in $\mathbf{M}(s)$ are now all causal transfer functions.

VIII. COMPARISION OF TFCM AND TWO COMMON MODEL FORMATS

The proposed linear TFCM model can be viewed as being a useful factorization of state space models:

$$\begin{aligned} \frac{d}{dt}\mathbf{x} &= \{\mathbf{A} + \mathbf{BX}(\mathbf{I} - \mathbf{DX})^{-1}\mathbf{C}\}\mathbf{x} \\ &+ \mathbf{B}\{\mathbf{X}(\mathbf{I} - \mathbf{DX})^{-1}\mathbf{D} + \mathbf{I}\}\mathbf{Uu} \\ \mathbf{y} &= \{\mathbf{Y}(\mathbf{I} - \mathbf{DX})^{-1}\mathbf{C}\}\mathbf{x} \\ &+ \{\mathbf{Y}(\mathbf{I} - \mathbf{DX})^{-1}\mathbf{DU} + \mathbf{W}\}\mathbf{u} \end{aligned} \qquad (26)$$

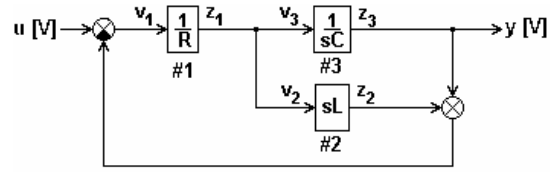


Fig. 4. Block Diagram for Simple R-L-C Circuit

where the dynamic model is written as a state space model:

$$\mathbf{M}(s) = \mathbf{C}(s\mathbf{I} + \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} \qquad (27)$$

(Equations 26 and 27 are derived from (A.1).)

The location of the TFCM connection matrices, $\mathbf{X}$, $\mathbf{U}$, $\mathbf{Y}$ and $\mathbf{W}$, within the model in (26) clearly shows the differences between the connection models of (3) and (4) and those of the proposed TFCM model of (5).

The linear TFCM model can also be viewed as a factorization of transfer function matrix models relating the system output variables, $\mathbf{y}$, to its input variables, $\mathbf{u}$:

$$\mathbf{G}(s) = \mathbf{YM}(s)[\mathbf{I} + \mathbf{XM}(s)]^{-1}\mathbf{U} + \mathbf{W} \qquad (28)$$

where the dynamic model could be a full matrix or a block diagonal matrix or an elemental diagonal model of SISO systems, depending on the level at which the analysis is aimed.

Clearly the TFCM provides a theoretical formulation of interconnected dynamic systems that differs from traditional state space and transfer function formulations.

IX. CONCLUSION

A mathematical model for interconnected dynamic systems has been proposed. It separates the system dynamics and interconnections into two set of equations – One defines the dynamics in terms of transfer functions; the other defines the connections as matrices of constants with elements that are real rather than binary or fractional. The transfer function models of its dynamic equation readily allow dead time terms and the equations of the TFCM model obey the traditional rules of matrix algebra.

Stability of the TFCM model is determined from its characteristic function and this is shown to be closely dependent on the structure of the characteristic connection matrix in the case of distributed systems. Interconnections between models alter the connection matrices but retain the dynamics of the original system. This is a significant feature of the proposed model and enables the analysis of connections independently of dynamics.

The use of the characteristic connection matrix in the TFCM model formulation is illustrated by two applications: In the first it guides the pairing of input-output variables in a twenty-seven model industrial plant. In the second it determines how non-causal dynamic models within a TFCM model can be inverted to form an equivalent model with causal dynamics.

REFERENCES

[1] Maciejowski, M (1989). *Multivariable feedback design.* Addison-Wesley, Wokingham.

[2] Guan, Z-H, G. Chen, X. Yu and Y. Qin. (2002). Robust decentralized stabilization for a class of large-scale time-delay uncertain impulsive dynamical systems. Automatica, 38, p2075-2084.

[3] Rosenbrock, H.H. (1974). Computer-aided control system design. Academic Press, London.

[4] Siljak, D.D. (1996). Decentralized control and computations: Status and prospects. A. Rev. Control, 20, 131-141.

[5] Callier, F.M., W.S. Chan and C.A. Desoer. (1978) Input-output stability of interconnected systems using decompositions. IEEE Trans. Automatic Control, AC-23 (2), 150-162.

[6] Groumpos, P.P and A.V. Pagalos. (1998) A two-level structural model for large scale systems. Computers in Industry, 36, 147-154.

[7] Guo Y, D.J. Hill and Y. Wang. (2000). Nonlinear decentralized control of large-scale power systems. Automatica, 36, 1275-1289.

[8] Hovd, M, R.D. Braatz and S. Skogestad (1997). SVD controllers for H_2-, H_∞- and μ- optimal control. Automatica, 33 (3), 433-439.

[9] Michel, A.N. (1983). On the status of stability of interconnected systems. IEEE Trans. Automatic Control, AC-28 (6), June 1983, p639-652.

[10] Siljak, D.D. (1991). Decentralized control of complex systems. Academic Press, Boston.

[11] VanAntwerp, J.G., A.P. Featherstone, R.D. Braatz. (2001). Robust cross-directional control of large scale sheet and film processes. J. Process Control, 11, 149-177.

[12] Zhisheng Duan, Jin-Zhi Wang and Lin Huang (2005). Input and output nonlinear systems. IEEE Transactions on Circuits and Systems, 52(3), 567-575.

[13] Braae, M. (2003). A connection theory for the analysis of large-scale systems. Proc. First African Control Conference, Cape Town, 3-5 December, p486-491.

[14] Dougherty, D and D. Cooper (2003). A practical multiple model adaptive strategy for single-loop MPC, Control Engineering Practice, 11(2), p141-159.

APPENDIX A. ALTERNATIVE TFCM FORMS

The TFCM model described in this paper (5) is linear in both its dynamics and its connections. This observation leads to the possibility of other formulations in which the dynamics and the connections are separable.

The transfer function matrix, $\mathbf{M}(s)$, of the TFCM model can be written as a state space model, $\mathbf{ABCD}$, to yield a state space form of the linear TFCM model:

$$\begin{aligned} \frac{d}{dt}\mathbf{x} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{v} \\ \mathbf{z} &= \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{v} \\ \mathbf{v} &= \mathbf{X}\mathbf{z} + \mathbf{U}\mathbf{u} \\ \mathbf{y} &= \mathbf{Y}\mathbf{z} + \mathbf{W}\mathbf{u} \end{aligned} \tag{A.1}$$

The TFCM model of (5) can also be generalized to include nonlinear dynamics and linear connections:

$$\begin{aligned} \frac{d}{dt}\mathbf{x} &= \mathbf{f}(\mathbf{x},\mathbf{v}) \\ \mathbf{z} &= \mathbf{g}(\mathbf{x},\mathbf{v}) \\ \mathbf{v} &= \mathbf{X}\mathbf{z} + \mathbf{U}\mathbf{u} \\ \mathbf{y} &= \mathbf{Y}\mathbf{z} + \mathbf{W}\mathbf{u} \end{aligned} \tag{A.2}$$

or to be linear in the dynamics with nonlinear connections:

$$\begin{aligned} \mathbf{z} &= \mathbf{M}(s)\mathbf{v} \\ \mathbf{v} &= \mathbf{X}(\mathbf{z},\mathbf{u}) \\ \mathbf{y} &= \mathbf{Y}(\mathbf{z},\mathbf{u}) \end{aligned} \tag{A.3}$$

or to be nonlinear in both its dynamics and its connections:

$$\begin{aligned} \frac{d}{dt}\mathbf{x} &= \mathbf{f}(\mathbf{x},\mathbf{v}) \\ \mathbf{z} &= \mathbf{g}(\mathbf{x},\mathbf{v}) \\ \mathbf{v} &= \mathbf{X}(\mathbf{z},\mathbf{u}) \\ \mathbf{y} &= \mathbf{Y}(\mathbf{z},\mathbf{u}) \end{aligned} \tag{A.3}$$

Of the formulations shown in the appendix only (A.1) has been used, specifically to illustrate the link between the models (3) or (4) and the TFCM model.

A High Performance Domain Specific OCR for Bangla Script

Md. Abul Hasnat S. M. Murtoza Habib Mumit Khan
Center for Research on Bangla Language Processing,
Department of Computer Science and Engineering,
BRAC University, 66 Mohakhali, Dhaka, Bangladesh
e-mail: mhasnat@gmail.com, murtoza@gmail.com, mumit@bracu.ac.bd

Abstract-Research on recognizing Bengali script has been started since mid 1980's. A variety of different techniques have been applied and the performance is examined. In this paper we present a high performance domain specific OCR for recognizing Bengali script. We select the training data set from the script of the specified domain. We choose Hidden Markov Model (HMM) for character classification due to its simple and straightforward way of representation. We examine the primary error types that mainly occurred at preprocessing level and carefully handled those errors by adding special error correcting module as a part of recognizer. Finally we added a dictionary and some error specific rules to correct the probable errors after the word formation is done. The entire technique significantly increases the performance of the OCR for a specific domain to a great extent.

I. INTRODUCTION

Methodically, character recognition is a subset of the pattern recognition area. However, it was character recognition that gave the incentives for making pattern recognition and image analysis matured fields of science [1]. In this literature we are considering the area of off-line character recognition. In a broad sense, from top level we can classify the area of Character Recognition into two divisions called machine printed and handwritten character recognition. So, these two categories can be termed as the two domains (group of documents that share similar layout structure) for Character Recognition. Now if we consider only the machine printed text document image with an additional specification of Bengali script then the number of domain will narrow down. We have observed that in Bengali scripts there exists a variety of documents that includes letters, text books, novels, official document, legacy document, newspapers, magazines, data entry form etc. From our experiment on different domain of documents we have seen that a universal technique with some common parameters may not fit perfect for all kinds of document. Based on this observation we feel that rather than a common technique with massive complexity and insignificant performance, we should choose a specific technique with adaptive parameters for a fixed domain or a set of domains to increase the performance significantly. This decision motivates us to think about domain specific solution for OCR. So, we choose our objective as to increase the performance of the OCR for a certain domain.

Research on OCR systems for recognizing Bangla characters have been started since mid 1980's, and a variety of different approaches were applied to examine the performance compared to the prior research results [2-13]. So far two implemented version of OCR for Bangla character recognition is reported [14, 15]. Among the research efforts, some were through the complete OCR system and the rest of the efforts were specific to the different level of OCR system like preprocessing, feature extraction, classification and post-processing. In the next section we will elaborately discuss about some of these research approaches those are relevant to our interest. The listed classifiers used in these research works are Nearest Neighbor Classifier [2], feature based tree classifier [3, 5, 7], template matching [5, 7], distance based classifier [8], Neural Network [9, 10, 12, 15] and Hidden Markov Model [13]. We have examined the performance of several classifier and we choose HMM (Hidden Markov Model) based classifier compare to others because of the following reasons:

- Simple and straightforward way of representation that provides the opportunity of dynamic and time imperceptible training at user end.
- Segmentation free approach.
- Shows significant performance for trained characters.

The performance of an OCR significantly depends on the performance of the recognizer. HMM (Hidden Markov Model) is a widely used classifier mostly for Speech and Handwriting recognition. Very few research works can be found on the usage of HMM for printed character recognition [13]. On the other hand we have reported very persuade performance for some domain specific HMM based recognizer [16, 17]. These considerations greatly motivate us to perform our research and development of an OCR for Bangla using HMM technique. We tested our implemented OCR for different types of document Images and recorded the results. The result shows that most of the errors occurred for the segmentation problem that happened due to the errors at multiple stages at preprocessing step. Based on our overall analysis we find out a specific data set for training. We performed error analysis and based on these errors we changed our training methodology and also placed an error correcting module at the end of basic character recognition result. After the basic word formation is done, we added a dictionary look-up based post-processor that provides up to a certain number of suggestions for the erroneous words. The combination of all these approaches leads to the increase of performance to a great extent. To the best of our knowledge

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 174–178.

this is the first reported attempt on domain specific OCR for recognizing Bangla text Image.

In this paper we will briefly present the complete methodology of the OCR technique with the probable errors encountered during recognition and also the solution of these errors using several techniques. In the rest of the paper at section 2 we briefly discuss about the related works, at section 3 we describe the methodology with several sub-sections, at section 4 we will perform result analysis and at last we end up with the conclusion.

II. Related Works

We briefly discuss some related work in this section. A great amount of work has been done by B. B. Chaudhuri and U. Pal since mid 1990's. Following them some other researchers have come up with a variety of innovative ideas. The relevant works are briefly discussed below.

Reference [3] described a complete OCR system for Bangla. A detail description of the characteristics of Bangla text is discussed here. They used a combination of template and feature matching approach for recognizing the character shapes. They used stroke features from each character and used a feature based tree classifier for character recognition. They classified the character set as basic and compound character. They used a simple dictionary lookup for OCR error correction.

Reference [4] used a technique for OCR error detection and correction on Bangla language. They used two separate lexicons of root word and suffixes. These errors are corrected by a fast dictionary access technique.

Reference [5] described their approach to recognize both Bangla and Devnagari scripts. At preprocessing level they applied Hough transform to find the skew angle, page layout analysis to handle multiple columns of text line and graphics, text line zoning technique for character segmentation. Unlike [3] they grouped characters into three classes named basic, modifier and compound character. They used feature based approach for basic and modifier character recognition and a combination of feature based template matching approach for the compound character recognition. However, recognition and error handling are almost similar as the approach described at [4].

Reference [7] presented a Complete Printed Bangla OCR System where they discussed about the difficulties encountered in Bengali script. In their approach the basic and modified characters are recognized by a structural-feature-based tree classifier and the compound characters are recognized by a tree classifier followed by template-matching approach. They used character unigram statistics to make the tree classifier efficient and several heuristics to speed up the template matching approach. A dictionary-based error-correction scheme has been used as a post processor.

Reference [8] give a brief overview of OCR research on Indian languages and also provide a substantial description of their work. They used a hybrid approach to recognize the parts of the conjunct that form part of a character class. To classify the segmented images into known classes, they used a set of filters and two distance based classifiers. They presented a two level partitioning scheme and search algorithm for the correction of optically read characters.

Reference [9] described their approach to recognize only the basic characters [3, 5]. They applied thinning and scaling at preprocessing level and multi-layered feed-forward back propagation neural network for character recognition purpose.

Reference [10] used curvature properties as local feature that is acquired from the slope distribution of chain code representation of each character. Their classification strategy was using Neural Network approach where for training and recognition conventional back propagation algorithm was performed.

Reference [12] presented a minimally segmented OCR where scaled the segmented image into a predetermined area. They extracted feature vector from a rectangular pixel map as a series of 0s and 1s. Finally they used a Kohonen neural network based classifier for character recognition.

The most recent work has been reported by M. A. Hasnat et al. [13] where they presented segmentation free OCR using Hidden Markov Model (HMM) based Recognizer. They applied Discrete Cosine Transform (DCT) to calculate the feature values. Their approach was to build separate model for each segmented primitive. Their approach shows significant performance for trained character.

So far we have seen two implementations of OCR for Bangla. They are BOCRA [14] and Apona Pathak [15]. BOCRA targets a very specific type of OCR problem, namely where the input images are high quality scans of high quality printed text written in a single font in a uniform point size. Apona Pathak has the ability to handle multiple font and size. However the performance of both applications suffers greatly from segmentation error.

III. Methodology

The procedural block diagram of the OCR system is shown in Fig. 1. This diagram is quite straightforward where each block can be further divided into several internal steps. We made our efforts in each subsection to minimize the errors based on our error analysis. The next sub-sections will elaborately describe each block.

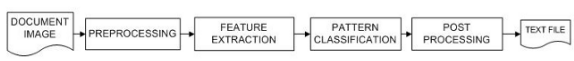


Fig. 1: Block Diagram of OCR system

A. Preprocessing

Preprocessing is the fundamental and very important stage. Lots of errors occurred at this step and for this reason enormous number of research work is going in this area which leads to a separate area of research called Document Image Analysis. Possible errors that may occur at this stage are briefly discussed at [1, 16]. During this research work we were aware of these errors and tried to overcome these errors at best. In this stage we perform the followings tasks:

1) Image Acquisition and Binarization: We used a flatbed scanner for Image acquisition and digitization. We are able to

process text images at any format. Then we applied traditional method to convert the color image to grayscale image if necessary. Next we perform thresholding operation to produce a binary image whose one state will indicate the printed text while the complementary state will correspond to the background. We experimented several traditional methods [18] and Otsu method [19] for thresholding and applied these methods based on domain type.

2) Noise elimination: We learned about different types of noise, their sources, and effects from the papers [1, 20]. From our observation we identified that for printed document the majority of the noises are the salt and pepper noises and the background noise. For background noise removal we used connected component information and eliminate the noise using statistical analysis. For other type of noise removal and smoothing we used wiener and median filters [21].

3) Skew detection and correction: We considered two methods [6, 11] for this purpose and we followed the approach discussed at [11]. First we identified the upper envelope and then we applied Radon transform to the upper envelope to get the skew angle. We applied generic rotation algorithm for skew correction and then applied bi-cubic interpolation.

4) Line, word and character level segmentation: We have studied several segmentation approaches discussed at [3, 5, 7-10, 12]. From implementation perspective we observed that, most of the errors occurred at character level segmentation. Line and word level segmentation failed due to the presence of noise which gives wrong estimation of the histogram projection profile. However character level segmentation mostly suffers from joining error (fail to establish a boundary where there should be one) and splitting error (mistakenly introduce a boundary where there should not be one). Considering all these we made our effort up to a minimal segmentation [12] and we resolved these issues during classification. Finally we used a simple technique similar to [3]. Fig. 2 shows the segmented units.

Fig. 3: Segmentation Result

The output from this block is a set of segmented image which is provided as an input to the next block where each one is processed separately.

B. Feature Extraction

At this stage we divide each segmented character image into several frames of fixed length (e.g. 8 pixels). Then we applied DCT (Discrete Cosine Transform) calculation over each pixel of the frames. We followed the similar technique described at [13]. The extracted feature for each character image is written into a file in a specific format. This process is shown in Fig. 3.

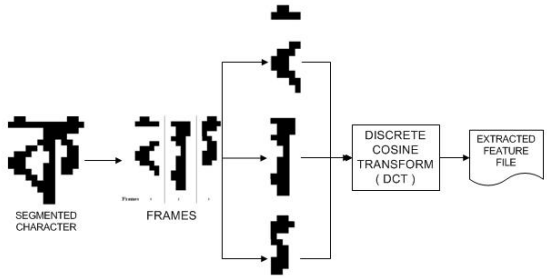


Fig 3: Feature Extraction process

C. Pattern Classification

This stage describes the training and recognition methodology. The extracted features for each segmented character are considered as the input for this stage. We followed almost similar strategy discussed at [13] for classification, however we did not limit ourselves on several issues like training from multiple samples and also the trained data representation using a fixed prototype model. We introduced the concept of dynamic training at any level of recognition and dynamic prototyping as well. For the recognition process we create a temporary model from the feature file of each character image and simply pass the model to the recognizer for classification. Like [13] we also used HTK recognizer [22] for our research and implementation.

1) Training: For training we create a separate model for each of the training character or symbol from the training data set. We estimated all around 650 training data unit (primitives and compounds) into the training data set based on our analysis on the OCR performance. This large amount of training data unit ensures the error tolerance at recognition. These samples are considered as the primitives for any trained OCR. We proposed dynamic training which enables us to train the OCR even after observing the recognition result and hence further improve the performance. We choose prototypes dynamically for the initialization of each model where each prototype contains the proper HMM model parameters like: Number of states, observation and transition matrix. HTK re-estimates the model parameters using this prototype and the extracted features.

Data Set for Training: In our training data set initially we considered only the alphabets of Bangla character set with the traditional segmentation method, but the recognition performance was not considerable. Then we added the compound characters into the training set and we obtain a good performance. However with this database the system was yet suffering from segmentation error occurred at the places of the vowel and consonant modifiers. So, finally we have taken the minimal segmentation approach [12] and added the characters with the vowel and consonant modifiers into the training set. During training, we must associate the appropriate Unicode character in the same order as they appear in the image.

2) Recognition: The recognition process is quite straightforward. The classifier temporary creates a model for

each minimally segmented character and recognizes this model using Viterbi decoder [13, 22].

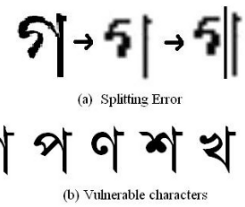


Fig. 4: (a) example of splitting error (b) vulnerable basic characters

We added a simple error correcting module at the end of basic character and symbol recognition that corrects the splitting errors which occurred due to the combination of over thresholding and segmentation problem. We identified that the characters that do not have full matra (baseline) over it, mostly suffers from this error. An example of the splitting error and also the list of vulnerable basic characters are shown in Fig. 4. We observed that the second part of each of these broken characters is classified as Bangla aakar (a vowel modifier) and the first part is misclassified or not classified. To solve this problem we trained the broken first part with special symbols and keep the second part as it is. After the recognizer classified the basic characters or symbols we resolved these erroneous issues in this module by a special table lookup. The output of this module is the words formed after the character level classification. Here we are considering this module as a part of our recognizer.

D. Post Processing

In this stage we used a suggestion based spelling checker for correcting the erroneously recognized words. We applied a technique based on the concept of the spelling checker proposed by Naushad et al. [23]. However instead of a phonetic encoding table we used a table that actually codes the graphitic symbols. We assign same code to those characters that are visually almost similar. Here we encode the characters based on the possible errors encountered at our observation on the OCR result without spell checker. In our approach rather than replacing the erroneous word we would like to provide a certain number of suggestions for that word.

IV. RESULT ANALYSIS

We tested the performance of the classifier in several domains with the specified training data set and we obtained an average of almost 98% accuracy of the classifier for properly binarized image and segmented characters. However from our analysis we are aware that once a split, join, or misalignment error is present in the output of the segmentation stage, even otherwise perfect classifiers will generally fail [16]. Our error correcting module and the postprocessor are capable to handle 14% - 17% of these errors depending on the different domains. Table I gives a clear scenario of the reported error, error tolerance rate and final accuracy at different level for different domain.

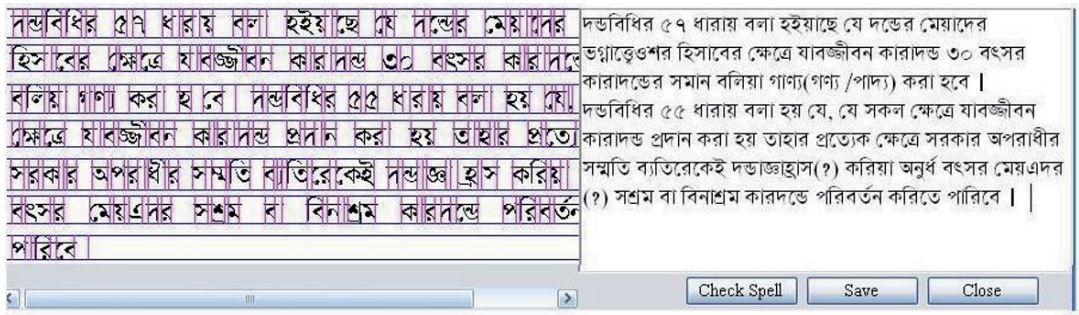


Fig. 5: a) Domain: Legal document

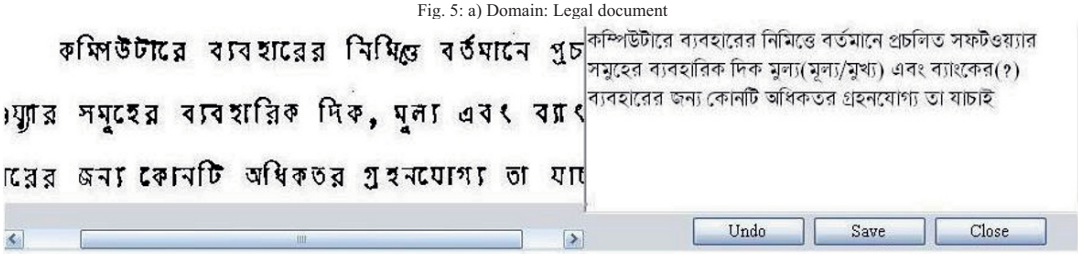


Fig. 5: b) Domain: Typewriting document

Fig. 5: Performance of the OCR for two different domain (a & b) document images.

TABLE I
LIST OF DOMAIN SPECIFIC PERFORMANCE

Domain Name	Classification Accuracy	Segmentation Error Rate	Total Error	Error Tolerance	Final Accuracy
Legal Document	98%	16%	18%	16%	98%
Typewriting	96%	15%	19%	16%	95%
Printed Article	97%	13%	16%	14%	98%

Fig. 5 shows the screen shot of the implemented version of our proposed OCR with the results.

V. CONCLUSION

This paper presents a complete Bangla OCR for domain specific document images. At different stages we tested several methods and choose the appropriate one for our purpose. We have done a complete analysis of the possible errors. Proper identification of the errors helps us to take right decisions to correct those errors at different stages. As a complete solution, the OCR shows high performance for specific domains. We have shown that at classification level we obtain massive accuracy, however segmentation problem degrades the accuracy and at the end we put our efforts to achieve the high accuracy by introducing an error correcting module with the recognizer and a suggestion based post processor.

VI. REFERENCE

[1] Line Eikvil, "Optical Character Recognition", "citeseer.ist.psu.edu/142042.html".

[2] A. K. Roy and B. Chatterjee, "Design of a Nearest Neighbor Classifier for Bengali Character Recognition", J. IETE, vol. 30, 1984.

[3] U. Pal and B. B. Chaudhuri, "OCR in Bangla: An Indo-Bangladeshi Language", Proc. of 12th Int. Conf. on Pattern Recognition, IEEE Computer Society Press, pp. 269-274, 1994.

[4] B. B. Chaudhuri and U. Pal, "OCR Error Detection and correction of an Inflectional Indian Language Script", Proceedings of ICPR, 1996.

[5] B. B. Chaudhuri and U. Pal, "An OCR System To Read Two Indian Language Scripts: Bangla And Devnagari (Hindi)", Proc. Fourth ICDAR, 1997.

[6] B.B. Chaudhuri and U. Pal, "Skew Angle Detection Of Digitized Indian Script Documents", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 19, pp.182-186, 1997.

[7] B. B. Chaudhuri and U. Pal, "A Complete Printed Bangla OCR System", Pattern Recognition, vol. 31, pp. 531-549, 1998.

[8] Veena Bansal and R.M.K. Sinha, A Devanagari OCR and A Brief Overview of OCR Research for Indian Scripts in Proceedings of STRANS01, held at IIT Kanpur, 2001.

[9] A. A. Chowdhury, Ejaj Ahmed, S. Ahmed, S. Hossain and C. M. Rahman, "Optical Character Recognition of Bangla Characters using neural network: A better approach". 2nd ICEE 2002, Khulna, Bangladesh.

[10] J. U. Mahmud, M. F. Raihan and C. M. Rahman, "A Complete OCR System for Continuous Bangla Characters", Proc. of the Conf. on Convergent Technologies, 2003.

[11] S. M. Murtoza Habib, Nawsher Ahmed Noor and Mumit Khan, Skew correction of Bangla script using Radon Transform, Proc. of 9th ICCIT, 2006.

[12] S. M. Shoeb Shatil and Mumit Khan, "Minimally Segmenting High Performance Bangla OCR using Kohonen Network", Proc. of 9th ICCIT, 2006.

[13] Md. Abul Hasnat, S. M. Murtoza Habib, and Mumit Khan, Segmentation free Bangla OCR using HMM: Training and Recognition, Proc. of 1st DCCA2007, Irbid, Jordan, 2007.

[14] http://bocra.sourceforge.net/doc/ last accessed Oct 22, 2007.

[15] http://www.apona-bd.com/apona-pathak/bangla-ocr-apona-pathak.html last accessed Oct 22, 2007.

[16] 16. A. Kornai and K. Mohiuddin and S. Connell, "An HMM-Based Legal Amount Field OCR System for Checks", 1995 IEEE International Conference on Systems, Man and Cybernetics, Vancouver BC, October 1995, 2800-2805.

[17] 17. A. Kornai, Experimental HMM-based postal OCR system, Proc. Int. Conf. Acoustics, Speech, Signal Processing, Munich, Germany, Vol. 4, 3177-3180, 1997.

[18] 18. John C. Russ, "The image processing handbook", CRC Press, Boca Raton, FL, USA, 1998.

[19] 19. N. Otsu, "A Threshold Selection Method from Gray-Level Histograms", IEEE Transactions on Systems, Man, and Cybernetics, 1979.

[20] 20. Yan Solihin and C.G. Leedham, "Noise and Background Removal from Handwriting Images", Proc. of the IASTED Int. Conf. on Intelligent Information Systems , 1997.

[21] 21. Tinku Acharya and Ajoy K. Ray, "Image Processing, Principles and Applications", John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.

[22] 22. The HTK Book available at http://htk.eng.cam.ac.uk/docs/docs.shtml

[23] 23. Naushad UzZaman and Mumit Khan, "A Double Metaphone Encoding for Bangla and its Application in Spelling Checker", Proc. 2005 IEEE Int. Conf. on Natural Language Processing and Knowledge Engineering, Wuhan, China, October 30 - November 1, 2005.

Tracking Performance of an Identical Master-Slave Teleoperation System Under Variable Time Delays

Mehmet Ismet Can Dede, and Sabri Tosunoglu

Florida International University
Department of Mechanical Engineering
10555 West Flagler Street
Miami, Florida 33174

***Abstract* – Teleoperation system performance suffers from the delays in communication lines. These delays are usually variable time delays. The customary wave variable technique used for constant time delays cannot provide satisfactory results to stabilize the system under variable delays. Another problem is observed in the position tracking performance of the limited-workspace teleoperation systems. Although wave variable technique is proven to track position demands successfully under constant time delays given that there is no communication failure, position tracking performance also degrades under variable delays. In this paper, a new controller is proposed to stabilize the system and enhance position tracking performance under variable time delays. This proposed controller is then applied to an identical master-slave teleoperation configuration which qualifies as a limited-workspace teleoperation system.**

I. Introduction

Teleoperation systems are composed of three subsystems: (1) a master, (2) a slave, and (3) a communication line. These subsystems work together to complete a task at sites that are either at distant locations and/or at places that are hazardous for humans. The human operator uses a master system to send out commands (or demands). The slave system is driven by these demands and the teleoperation system sends back sensory information to the master. This type of teleoperation is named bilateral. One special type of bilateral teleoperation is force-reflecting bilateral teleoperation. The slave sends back force information to drive the actuators of the master system. This enables the human operator to sense the working environment of the slave better. The types of teleoperation are briefly described in the next section.

Teleoperation has attracted interest in researchers partly because it is utilized in a wide variety of applications that span the space, nuclear reactors, battlefields, security needs, undersea tasks, medical operations, and training, among others. Often teleoperation has been associated with the time delay problem when it is implemented over longer distances (when controlling a device on Mars, for instance), or when it utilizes the Internet infrastructure to communicate with a remote system. The system and its control become more complicated when force reflection is introduced. Such a system used under significant time delays poses a challenging controller design problem since the stability of such a system cannot be easily guaranteed.

Many researchers have studied this problem, but Anderson and Spong were perhaps the first to use the wave variable method to control bilateral teleoperation [1]. Also, Niemeyer and Slotine [2], and Munir and Book [3] have implemented this method to teleoperation systems.

Although the wave variable technique guarantees stability under constant time delays, it introduces offsets in position tracking if the teleoperation system experiences a communication loss. This was studied earlier and a position feedforward component was presented in [4]. Position tracking performance is the priority in limited-workspace teleoperation systems. In unlimited-workspace teleoperation systems the tracking priority is usually given to the velocity. Therefore, the addition of the position feedforward component to the customary wave variable technique is not necessary for these systems.

The customary wave variable technique fails to stabilize the system under variable time delays in both limited- and unlimited-workspace teleoperation. An adaptive gain was introduced in [5]. This component guarantees the stability of the teleoperation system while it cannot compensate for position offsets if it is used as the only additional component for the limited-workspace teleoperation with variable time delays.

The experiments for the existing and the modified algorithms are conducted by using an experimental setup. This setup consists of a gimbal-based joystick as the master and the virtual model of it as the identical slave system. A more detailed description of the experimental setup is presented later in the paper.

The next section provides an overview of teleoperation systems. Later, the original wave variable controller and the proposed modifications are presented. Performance of the proposed controller is evaluated experimentally and the results are provided.

II. Teleoperation Overview

A teleoperation system is generally used in one of two conditions: One condition is when it is necessary to accomplish a task at a distant site from the operator, and the other condition is when the task is carried out in an environment hazardous for humans. In both cases, the human operator is placed at the other end of the teleoperation system, sending signals to control the slave robot via the master

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 179–184.

system (joystick). In other words, the slave robot controlled by the human operator takes place of the human that is expected to work on the task. This substantially reduces the risk to humans and the costs associated in manned missions.

There are many applications of the teleoperation systems. For instance, Japanese researchers have studied the ground-space telerobotics [6]. The study by Cavusoglu [7] involves transforming a surgical robot for human telesurgery. Sitti [8] investigates teleoperated nanomanipulation. There are also numerous examples for military, hazardous environment and undersea teleoperations.

Telepresence can be considered as a measure of the quality of a teleoperation experience. Ideally, the information from the remote environment (visual, aural, haptic, etc.) is displayed in such a way that the operator "feels" as if he/she is actually present at the remote environment. Teleoperation systems can be branched into two types considering the concept of telepresence.

In unilateral teleoperation, the slave robot does not send back any sensory information to the master. Hence, the telepresence concept does not exist in this type of teleoperation. In contrast, bilateral teleoperation calls for the slave to send feedback. This means that the information flow is in both directions.

One of the most common bilateral teleoperation is called force-reflecting bilateral teleoperation. Slave reflects back the interaction forces to the master side. The actuators of the master are then driven by the sensory information to make the human operator feel the slave environment. Researchers in general agree that having force-reflection accompanied with visual feedback provides sufficient telepresence for most of the teleoperation applications [9].

Bilateral teleoperation systems can be further classified as limited- and unlimited-workspace teleoperation. Teleoperation systems using serial or parallel slave manipulators with limited workspace are defined as limited-workspace teleoperation. Telemanipulation of an industrial robot arm is an example to this type of teleoperation.

Teleoperation systems composed of a mobile platform or any unlimited-workspace slave is referred as unlimited-workspace teleoperation. Telemanipulation of any mobile robotic system whether it operates on ground, water or in air is categorized as unlimited-workspace teleoperation.

It is noted that the motion mapping of the two teleoperation types are not the same in practice. In limited-workspace teleoperation, generally master position and orientation information is mapped into the Cartesian position and orientation of the end-effector. In unlimited-workspace teleoperation, the position information from the master is generally mapped as velocity demand for the end-effector of the slave. Therefore, position tracking becomes a priority for the limited-workspace case while velocity tracking for unlimited-workspace teleoperation.

III. Wave Variable Technique

The wave variable technique in terms of the scattering transformation – a mapping between the velocity and force signals, and the wave variables [2] – is illustrated in Fig. 1.

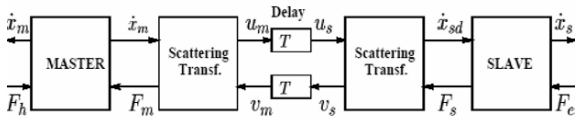


Fig. 1. Scattering transformation for teleoperation with constant time delay.

This transformation is described as follows [3]:

$$u_s = \frac{1}{\sqrt{2b}}(b\dot{x}_{sd} + F_s); \quad u_m = \frac{1}{\sqrt{2b}}(b\dot{x}_m + F_m)$$
$$v_s = \frac{1}{\sqrt{2b}}(b\dot{x}_{sd} - F_s); \quad v_m = \frac{1}{\sqrt{2b}}(b\dot{x}_m - F_m) \tag{1}$$

where $\dot{x}_m$ and $\dot{x}_s$ are the respective velocities of the master and the slave. F_h is the torque applied by the operator, and F_e is the torque applied externally on the remote system. F_m is the force reflected back to the master from the slave robot. F_s is the force information sent from the slave to master. $\dot{x}_{sd}$ is the velocity derived from the scattering transformation at the slave side. The wave variables are defined by u and v. The wave impedance in (1) is denoted as b.

The power, P_{in}, entering a system can be defined as the scalar product between the input vector x and the output vector y. Such a system is defined to be passive if and only if the following holds:

$$\int_0^t P_{in}(t)d\tau = \int_0^t x^T y d\tau \geq E_{store}(t) - E_{store}(0) \tag{2}$$

where $E(t)$ is the energy stored at time t and $E(0)$ is the initially stored energy. The power into the communication block at any time is given by

$$P_{in}(t) = \dot{x}_{md}(t)F_m(t) - \dot{x}_{sd}(t)F_s(t) \tag{3}$$

In the case of the constant communications delay where the time delay T is constant,

$$u_s(t) = u_m(t-T); \quad v_m(t) = v_s(t-T) \tag{4}$$

Substituting these equations into (3), and assuming that the initial energy is zero, the total energy E stored in

communications during the signal transmission between master and slave is found as

$$E = \int_0^t P_{in}(\tau)d\tau = \int_0^t (\dot{x}_{md}(\tau)F_m(\tau) - \dot{x}_{sd}(\tau)F_s(\tau))d\tau$$
$$= \frac{1}{2}\int_0^t (u_m^T(\tau)u_m(\tau) - v_m^T(\tau)v_m(\tau) + v_s^T(\tau)v_s(\tau) - u_s^T(\tau)u_s(\tau))d\tau$$
$$= \frac{1}{2}\int_{t-T}^t (u_m^T(\tau)u_m(\tau) + v_s^T(\tau)v_s(\tau))d\tau \geq 0 \quad (5)$$

Therefore, the system is passive independent of the magnitude of the delay *T*. In other words, the time delay does not produce energy if the wave variable technique is used. Therefore, it guarantees stability for the time-delayed teleoperation given that the system is stable when there are no time delays.

The customary formation of the wave variable technique is illustrated in Fig. 1. The controller on the slave side is usually a velocity controller. A representation of the basic velocity controller block is shown in Fig. 2. The velocity error is calculated in Cartesian space and then translated into joint velocity errors using the inverse of the Jacobian matrix. A general type of proportional-derivative control is applied to calculate the driving torque for each joint. The *N* in the block diagram of Fig. 2 represents the feed-forward torque input to counteract the centrifugal, Coriolis and gravitational forces.

Calculation of the errors in Cartesian space enables the usage of unlike masters and slaves. An example of such teleoperation is given in this paper as well as an example for the identical subsystem teleoperation.

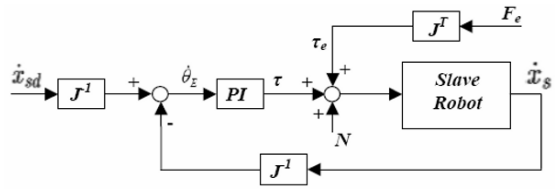


Fig. 2. Customary velocity control of the slave manipulator.

The system dynamics is written for teleoperation with the customary wave variable control as

$$M_m\ddot{x}_m + B_m\dot{x}_m = F_h - F_m\,; \quad M_s\ddot{x}_s + B_{s1}\dot{x}_s = F_s - F_e \quad (6)$$

where

$$F_s(t) = K_I\int_0^t \left(J^{-1}\dot{x}_{sd} - J^{-1}\dot{x}_s\right)ds + K_v\left(J^{-1}\dot{x}_{sd} - J^{-1}\dot{x}_s\right)$$

This controller produces acceptable system response when the wave variable technique is active but the communication is never lost between the master and the slave. When the communication is lost for limited periods, a drift between the master and slave position tracking occurs.

A. Position Feedforward Component

The authors propose a feedforward position demand to modify the wave variable technique. This demand is sent from the master system directly to the slave without integrating it in the scattering transform. This modification does not include a force feedforward component as in [10] because no drifts have been observed between the slave and the master force information in experimental studies [4]. The block diagram of the proposed algorithm is given in Fig. 3.

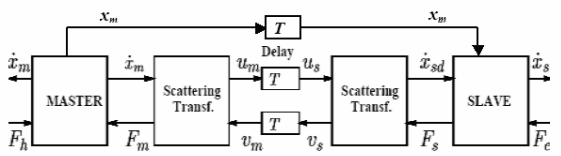


Fig. 3. Modification to wave variable block diagram.

The slave controller block diagram is also modified to comply with the new setting of the wave variable technique. As observed in Fig. 4, the position error is calculated in the joint space. The motion demand from the master received in Cartesian space is transformed into the joint space by using the inverse of the Jacobian, *J*, and the inverse kinematics, *IK*. Later, the demand in joint space is compared to the joint sensor readings to form joint motion errors. This type of controller is of course feasible for those manipulators for which the inverse kinematics solutions are easy to obtain. Fortunately, almost all of the industrial manipulators are of this kind [11, 12].

After the modification to the wave variable technique, the system dynamics is written as

$$M_m\ddot{x}_m + B_m\dot{x}_m = F_h - F_m\,; \quad M_s\ddot{x}_s + B_{s1}\dot{x}_s = F_s - F_e \quad (7)$$

where

$$F_s = K_d\left(J^{-1}\dot{x}_{sd} - J^{-1}\dot{x}_s\right) + K_p\left(IK(x_m(t-T)) - IK(x_s)\right)$$

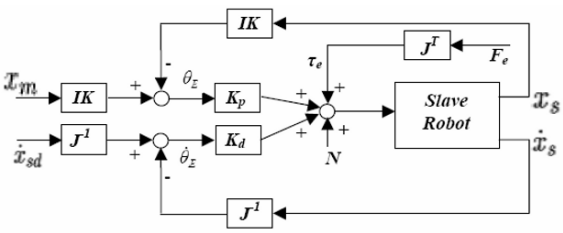


Fig. 4. Modified slave controller.

The control gains used in customary wave variable slave controller may be applied to the modified version. Therefore

the modified controller's gain magnitudes of K_d and K_p may be selected as equal to the magnitude of K_v and K_I.

B. Adaptive Gain Component

The block diagram in Figure 5 shows the modification described in [5] for the wave variable method for the variable time-delayed teleoperation.

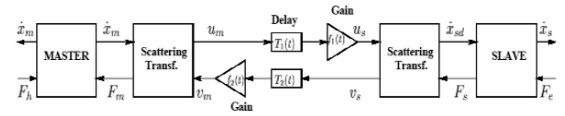


Fig. 5. Adaptive wave variable technique for teleoperation with variable time delays [5]

Time varying delay modifies the transmission equations to:

$$u_s(t) = u_m(t - T_1(t)); \quad v_m(t) = v_s(t - T_2(t)) \tag{8}$$

where, $T_1(t)$ is the variable time delay in the path from the master to the slave and $T_2(t)$ is the variable time delay in the path from the slave to the master. In [5], it is assumed that the frequency of change in time delays remain limited:

$$\frac{dT_i}{d\tau} < 1; \quad i = 1, 2 \tag{9}$$

Substituting the modified transmission equations to the equation for the total energy stored in the communication line (5)

$$E = \frac{1}{2}\Bigg[\int_{-T_1(t)}^{} u_m^T(\tau)u_m(\tau)d\tau + \int_{-T_2(t)}^{} v_s^T(\tau)v_s(\tau)d\tau - \int_{0}^{-T_1(t)} \frac{T_1'(\sigma)}{1-T_1'(\sigma)} u_m^T(\sigma)u_m(\sigma)d\sigma - \int_{0}^{-T_2(t)} \frac{T_2'(\sigma)}{1-T_2'(\sigma)} v_s^T(\sigma)v_s(\sigma)d\sigma \Bigg] \tag{10}$$

where $\sigma = \tau - T_i(\tau) = g_i(\tau)$ and $T_i'(\sigma) = \dfrac{dT_i}{d\tau}\Big|_{\tau = g_i^{-1}(\sigma)}$.

The last two terms in (10) show that passivity can not be guaranteed for variable time-delayed teleoperation. In the modified wave variable method, shown in Fig. 5, a time varying gain f_i is inserted after the time varying delay block. Therefore, the new transmission equation becomes:

$$u_s(t) = f_1(t)u_m(t - T_1(t)); \quad v_m(t) = f_2(t)v_s(t - T_2(t)) \tag{11}$$

The total energy stored can be re-written using the new transmission equations as

$$E = \frac{1}{2}\Bigg[\int_{-T_1(t)}^{} u_m^T(\tau)u_m(\tau)d\tau + \int_{-T_2(t)}^{} v_s^T(\tau)v_s(\tau)d\tau - \int_{0}^{-T_1(t)} \frac{1 - T_1'(\sigma) - f_1^2}{1 - T_1'(\sigma)} u_m^T(\sigma)u_m(\sigma)d\sigma - \int_{0}^{-T_2(t)} \frac{1 - T_2'(\sigma) - f_2^2}{1 - T_2'(\sigma)} v_s^T(\sigma)v_s(\sigma)d\sigma \Bigg] \tag{12}$$

If f_i is selected that $f_i^2 = 1 - T_i'$ for the total energy stored equation, the last two terms of the equation are eliminated and it can be said that the system is passive. In fact, the variable time-delayed system is considered to be passive if f_i satisfies the following condition [5]:

$$f_i^2 \le 1 - \frac{dT_i}{dt}; \; i = 1, 2 \tag{13}$$

This variable gain is called adaptive since it adapts itself with respect to the change in time delays.

IV. Experimental Setup

A two-degree-of-freedom (DOF) gimbal-based joystick is used as the master [13]. The joystick has uncoupled motions about the two axes due to its gimbal-based design. Fig. 6 shows the master joystick with its virtual replica.

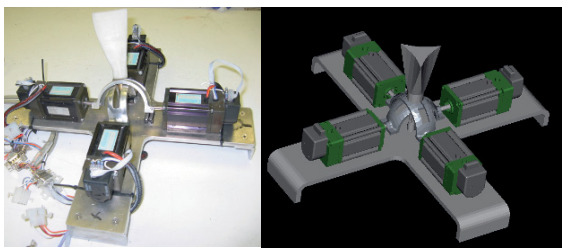


Fig. 6. Two-DOF master joystick (l) its virtual replica (r)

Each joint is designed to be bedded in between to two servomotors. Hence, joint level fault tolerance is achieved by connecting two servomotors to each link. If one servomotor fails, the other servomotor coupled to the same joint is able to take over the task. Each servomotor has an encoder connected to the rear end of its shaft. The demands of the human operator are sensed through these encoders. The measured change in the position of the joystick is then transmitted to the slave as manipulation demands.

The integration of the two subsystems of teleoperation is accomplished through the Matlab Simulink environment. Galil motion control card (DMC 18x2) is used to drive the servomotor on the actual joystick and also to gather the encoder readings from each axis. A driver is developed as an interface in simulation environment to send and receive data with the control card as detailed in [13]. The simulation is then synchronized with the real-time clock to run the real-time

tests. The sampling rate is selected to be 100 Hz. It is an acceptable rate for testing a system that experiences time delays that range from 0.1 to 0.5-second.

The slave is constructed as virtual replica of the master system. This indicates that the slave has the same workspace with the master and does not require mapping between the master and the slave motions. The master joint motion readings are directly sent as demands to the joint actuators of the slave.

This experimental setup requires virtual modelling of the robot. The virtual rapid robot prototyping concept presented by Dede and Tosunoglu [14] is used to construct the robots in virtual environment. First, the manipulator is constructed in a computer-aided-design software environment. Then, the material, inertial and mechanism parameters are translated into the Matlab© environment.

V. Experimental Results

The position feedforward component is only used for the y-axis controller. The adaptive gain is applied for both axes as the system experiences variable time delays. The adaptive gain is switched off through the end of the experiments to observe the performance of the customary wave variable technique under variable time delays. Since both axes are uncoupled, the motion along one axis does not affect the motion along the other axis. Hence, comparison of both controllers becomes possible in one experiment. Fig. 7 shows the time delays during the experiments and the change in the gain as it adapts itself to the variations in time delays. The magnitude of the time delays is consistent with the internet delays given in [3].

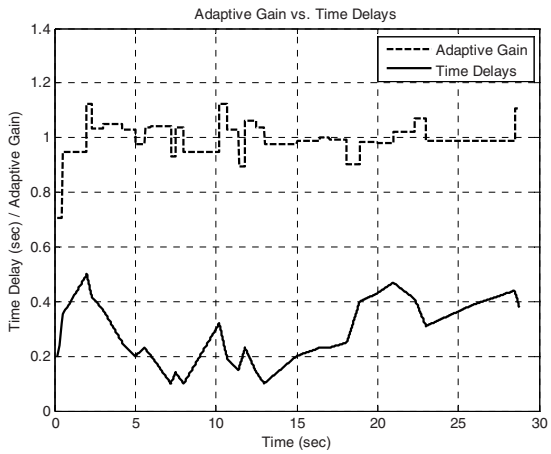


Fig. 7. Changes in adaptive gain due to the time delays in identical master-slave teleoperation.

The first set of figures represents the performance of the controller for the x-axis. The controller for this axis is the customary wave variables with the adaptive gain.

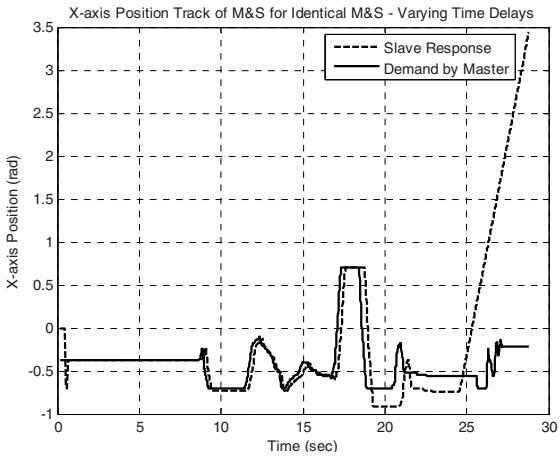


Fig. 8. Position tracking performance of the wave variable technique with adaptive gain under variable time delays

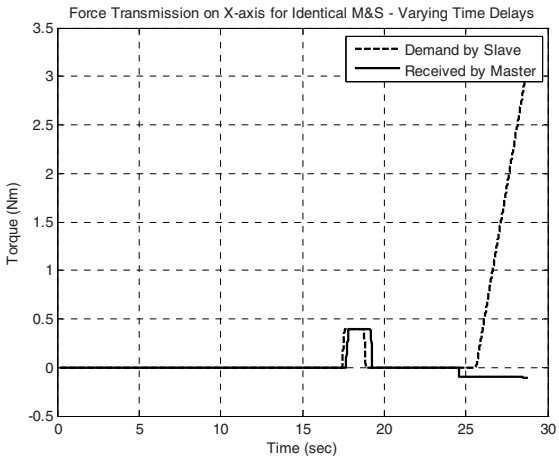


Fig. 9. Force tracking performance of the wave variable technique with adaptive gain under variable time delays

The adaptive gain is switched off after the 24th second in the experiments. As it is observed from Fig. 8 and 9, the system becomes unstable and the master cannot control the motions of the slave. It is also observed from Fig. 8 that the slave lost the position tracking of the master. Although the system does not experience any communication losses, a drift between the positions of the master and the slave is formed as a result of the variable time delays. No offset is observed in force tracking performance. Therefore, position feedforward component is employed for the controller on the y-axis. The following figures illustrate the results of the variable time delay experiments on the y-axis motion of the joystick.

The slave does not lose track of the position when the position feedforward term is added to the controller with adaptive gain. This is clearly observed in Fig. 10 and 11. The slave position does not even diverge from the demand after

the adaptive gain is switched off after the 24th second. Instead, an offset is observed in Fig. 10. Although the motion of the slave is kept stable with this controller without the adaptive gain, the force tracking is not possible and the system becomes unstable.

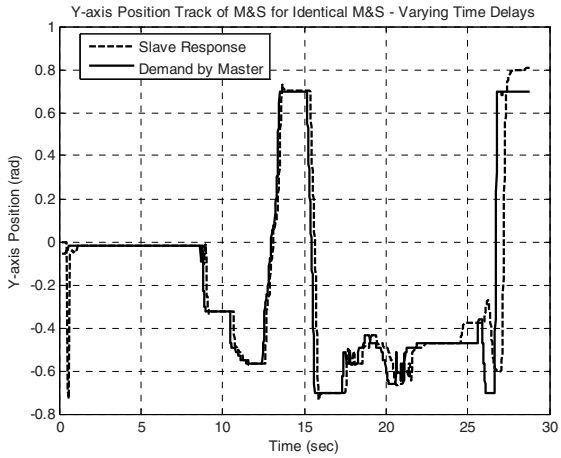


Fig. 10. Position tracking performance of the wave variable technique with adaptive gain and position feedforward components under variable time delays

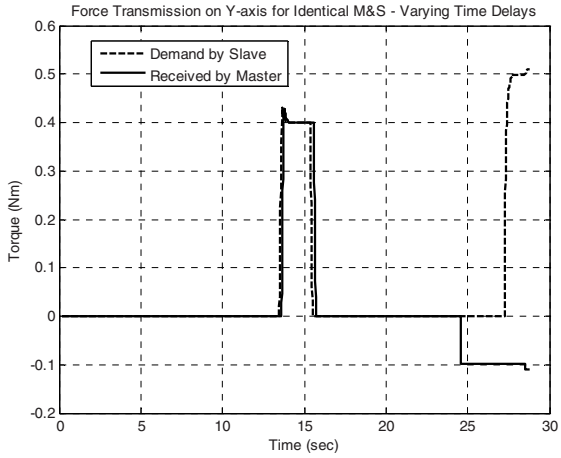


Fig. 11. Force tracking performance of the wave variable technique with adaptive gain and position feedforward components under variable time delays (y-axis)

VI. Conclusions

The control algorithms developed for the limited-workspace teleoperation systems experiencing constant time delays cannot guarantee system stability under variable time delays. An adaptive gain component is introduced to be utilized in such cases. As a result of the addition of this component, the limited-workspace teleoperation system under variable time delays is stabilized. In the experiments, it is also shown that when the adaptive gain becomes inactive the system eventually becomes unstable. It is also concluded that the position feedforward component is necessary to keep the system stable and to track the position demands under variable time delays.

The experiments indicate that the wave variable technique with adaptive gain and position feedforward component is the most suitable controller for identical master-slave teleoperation systems experiencing variable time delays.

Extension of this approach to non-identical master-slave teleoperation systems as well as to unlimited-workspace teleoperation systems is under investigation.

References

[1] R. J. Anderson, and W. Spong, "Bilateral Control of Teleoperation with Time Delay," in IEEE Transactions on Automation and Control, Vol. 34, Is. 5, pp. 494-501 1989.
[2] G. Niemeyer, and J. Slotine, "Using Wave Variables for System Analysis & Robot Control," in Proceedings of the IEEE International Conference on Robotics and Automation, Albuquerque, NM, April 1997.
[3] S. Munir, "Internet-Based Teleoperation," Ph.D. Dissertation, Georgia Institute of Technology, Atlanta, GA, 2001.
[4] M..I.C. Dede, and S. Tosunoglu, "Modification of the Wave Variable Technique for Teleoperation Systems Experiencing Communication Loss," in Proceedings of the 7th IEEE International Symposium on Computational Intelligence in Robotics and Automation, pp. 380-385, Jacksonville, Florida, June 20-23, 2007.
[5] N.Chopra, M.W. Spong, S. Hirche, and M. Buss, "Bilateral Teleoperation Over The Internet: The Time Varying Delay Problem," in Proceedings of the American Control Conference, vol. 1, pp. 155-160, Denver, CO, June 4-6, 2003.
[6] P. Huang, Z. Liu, G. Zhao, W. Xu, and B. Liang, "A Ground Teleoperation Experimental System of Space Robot using Hybrid Approach," Proceedings of the IEEE International Conference on Integration Technology, pp. 593-598, Shenzhen, China, march 20-24, 2007.
[7] M.C. Cavusoglu, "Telesurgery and Surgical Simulation: Design Modeling, and Evaluation of Haptic Interfaces to Real and Virtual Surgical Environments," Ph.D. Dissertation, UC at Berkeley, Berkeley, California, 2000.
[8] M. Sitti, "Teleoperated and Automatic Control of Nanomanipulation Systems using Atomic Force Microscope Probes," in Proceedings of the IEEE Conference on Decision and Control, Maui, Hawaii, December 2003.
[9] A.N. Tanner, and G. Niemeyer, "High-Frequency Acceleration Feedback in Wave Variable Telerobotics," in IEEE/ASME Transaction on Mechatronics, Vol. 11, Is. 2, pp. 119-127, April 2006.
[10] N. Chopra, M.W. Spong, R. Ortega, and N.E. Barbanov, "On Tracking Performance in Bilateral Teleoperation", in IEEE Transaction on Robotics, Vol. 22, Is. 4, pp. 861-866, 2006.
[11] M.K. Özgoren, "Topological Analysis of 6-Joint Serial Manipulators and Their Inverse Kinematics Solutions," in Mechanism and Machine Theory, Vol. 37, Is. 5, pp. 511-547, 2002.
[12] J. Duffy, "Analysis of Mechanisms And Robot Manipulators" Wiley, New York, NY, 1980.
[13] M.I.C. Dede, and S. Tosunoglu, "Development of a Real-Time Force-Reflecting Teleoperation System Based on Matlab© Simulations," in Proceedings of the 19th Florida Conference on Recent Advances in Robotics, Miami, Florida, May 25-26, 2006.
[14] M.I.C. Dede, and S. Tosunoglu, "Virtual Rapid Robot Prototyping," in ASME Early Career Technical Journal, Vol. 5, Is. 1, pp. 7.1-7.8, 2006.

New Concept in Optimizing Manipulability Index of Serial Manipulators, Using SVD Method

Mohammed Magdy[1], Ayssam ElKady[2,] and Tarek Sobh [2]

1- Mechanical Engineering Department
2-Computer Science Department
University of Bridgeport, Bridgeport, CT 06604 USA
{Mohammem, Sobh, aelkady}@bridgeport.edu
2007

Abstract-**Studying the Manipulability index for every point within the workspace of any serial manipulator is considered one of the important problems, required for designing trajectories or avoiding singular configurations. We had to go through this problem for our project of optimizing D-H parameters of extending link/s for Mitsubishi which is existed at our lab. First step is to get all points within the workspace envelope with different manipulability bands and to find most the high band to start forward kinematic design and optimization of additional link/s parameters**.

We have discovered a new concept in this paper ,related for one method used before in measuring kinematic Manipulability .We have made simulations supporting our method on Puma560, Mitsubishi movemaster® , and six degrees of freedom manipulator.

I.Introduction

The benefit of using quantitative measure is well known in engineering systems. In other words, a quantitative measure provides one with a rational base which one can - without depending on experience -analyze, design and control the systems as follows: [1]

1-It helps one to evaluate performance of the system in simulation mode before involving in real time mode.

2-It helps one to design new system or at least modifying existing system from points or level providing high performance from this quantitative measure's view.σ

Manipulability measure is considered one of the most important quantitative measure in Robotic field .many mathematical methods have been proposed to describe robot's dexterity (Manipulability) and whether it can be optimized by observing the distance between singular configuration or not. Most of these previous researches have been tested while the manipulator is passing through line segment as robot's path from the basis of manipulator until last point on the robot's path, that manipulator can reach (at the boundary of the workspace envelope). The first points on path considered one with low dexterity (Manipulability) or even zero dexterity because in this configuration, one or more links are completely folded or at least are very close to this singular configuration. Also, when the manipulator reaches last point in the line segment, it also reaches another singular configuration, while it is fully extended.

Many mathematical formulations have been proposed to express the Manipulability measure; the product of all singular values of Jacobian, det $(JJ^T)^{1/2}$, by Yoshikawa. The minimum singular value of the Jacobian, min (σ (J)), by Klein and Blaho as a measure of the closeness of the manipulator postures to singularities. The reciprocal of the condition number, termed the conditioning index (CI), has been proposed by Angeles to measure the closeness of a configuration to singularity.[2]

As mentioned in [3], the determinant of jacobian can not be used for expressing Manipulability's index .It just becomes zero when a manipulator reaches any singular configuration. Other method has been proposed called reciprocal of the Jacobian as in [1,2,3] and we have tried to test this method with different concept and did not prove any efficiency for non linear path.

It is worth to mention that we are interested in arm Manipulability(that can be measure as distance from singularity) not manipulability of orientation, so we also tried to decouple the jacobian matrix according to [6] and the result , unfortunately was not satisfied at all ; bands were intersected with each others. Last method proposed was Singular Value Decomposition of jacobian matrix.

II.Singular Value Decomposition: [5]

Singular value decomposition is a mathematical technique to optimize the singular form of the any square or non square matrix $A_{(m \times n)}$, depending on eigen values, this is technique has been used in many engineering application.

$$A_{m \times n} = U_{m \times m} S_{m \times n} V^T_{n \times n} \quad (1)$$

$$U^T U = I \quad (2)$$

$$V^T V = I \quad \text{(i.e. U and V are orthogonal)} \quad (3)$$

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 185–191.

$$S=\begin{pmatrix}\sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0\\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & 0 & 0 & \cdots & 0\\ 0 & 0 & \cdots & \sigma_m & 0 & \cdots & 0\end{pmatrix} \quad (4)$$

We have constructed the jacobian according to [5]

$$J=[J_1\ J_2\ldots\ldots\ldots\ J_n]; \quad (5)$$

Where n is number of degrees of freedom

$$J_i=\begin{pmatrix}Z_{i-1}x(O_n-O_{i-1})\\ Z_{i-1}\end{pmatrix} \quad \text{For revolute joint (6)}$$

U and V are orthogonal matrix means U x U^T =1 & V x V^T = 1 and the singular values are in Descending orders $\sigma_1 \geq \sigma_2 \geq \ldots\ldots \geq \sigma_m$. Matrix A has full rank means rank of A= m when $\sigma_m \neq 0$. When $\sigma_m \approx 0$ with small tolerance the rank of matrix A looses one or more degrees of freedom and that means mathematically, the matrix looses one or more degree of freedom (one row or column becomes non independent ; it can be optimized by multiply another row or column by factor). Physically, any manipulator has two joint's axes coinciding on each other. For example, puma 560® at initial position; the sixth joint and fourth joint are on same axes so if the fourth or sixth joint moves, the result motion (orientation) will be the same. That means the manipulator looses one degree of freedom.

In past research, there was an argument about whether the minimum value of theses σs or the multiplication of all σs exactly represent the Manipulability's index.[4]

In this paper, we will propose a new concept for measure and observe this index plus make sure by drawing the bands of this index. Moreover, it will be discussed a new relationship between the minimum rank of the jacobian matrix, the manipulator can reach in any singular configuration (assuming robot's joint from [0 to 360 degree]), and the order of one of these σs that can exactly express the Manipulability's index.

III. NUMERICAL EXAMPLE

In case of singular configuration of Puma 560 at Q= [0, 0,-90, 0, 0, 0]

$$J=\begin{bmatrix}0 & 0 & 0 & 0 & 0 & 0\\ 20 & 0 & 0 & 0 & 0 & 0\\ 0 & 20 & 10 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 1\\ 0 & 1 & 1 & 0 & 1 & 0\\ 1 & 0 & 0 & 0 & 0 & 0\end{bmatrix}$$

$$U=\begin{bmatrix}0 & 0 & 0 & 0 & 1 & -0.0034\\ 0 & -0.9988 & 0 & 0 & -0.0002 & -0.0499\\ -0.9982 & 0 & 0 & 0.0600 & 0 & 0\\ 0 & 0 & -1 & 0 & 0 & 0\\ 0.06 & 0 & 0 & 0.9982 & 0 & 0\\ 0 & -0.0499 & 0 & 0 & 0.0034 & 0.9987\end{bmatrix}$$

$$S=\begin{bmatrix}22.4010 & 0 & 0 & 0 & 0 & 0\\ 0 & 20.0250 & 0 & 0 & 0 & 0\\ 0 & 0 & 1.4142 & 0 & 0 & 0\\ 0 & 0 & 0 & 1.0935 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix}$$

$$V=\begin{bmatrix}0 & -1.0000 & 0 & 0 & 0 & 0\\ -0.8939 & 0 & 0 & 0.1852 & 0.4074 & -0.0270\\ -0.4483 & 0 & 0 & -0.3638 & -0.8147 & 0.0539\\ 0 & 0 & -0.7071 & 0 & -0.0467 & -0.7056\\ -0.0027 & 0 & 0 & -0.9129 & 0.4074 & -0.0270\\ 0 & 0 & -0.7071 & 0 & 0.0467 & 0.7056\end{bmatrix}$$

It is obvious that in singular matrix S; σ_5 and σ_6 are zero with small tolerance .That is because there are two singular cases in its Configuration; the forth and sixth joint are on same axis plus it is in singular arm configuration , so also σ_5 is zero .

TABLE 1
D-H Parameters of Puma 560[7]

i	α	A	Θ	D	Initial limit	Final limit	Joint's Type
1	90	0	*	0	-170	170	R
2	0	0.431800	*	0	-225	45	R
3	-90	0.020300	*	0.150050	-250	75	R
4	90	0	*	0.431800	-135	100	R
5	-90	0	*	0	-100	100	R
6	0	0	*	0	-180	180	R

TABLE 2
D-H Parameters of Six Degrees of freedom Serial Manipulator

i	α	A	Θ	D	Initial limit	Final limit	Joint's Type
1	90	0	*	10	Same joint's limit of Puma560		R
2	0	10	*	0			R
3	-90	0	*	0			R
4	90	0	*	10			R
5	-90	0	*	0			R
6	0	0	*	0			R

TABLE 3

D-H Parameters of Mitsubishi Movemaster®

i	α	A	Θ	D	Initial Limit	Final Limit	Joint's Type
1	90	0	*	300	-150	150	R(Pos.)
2	0	250	*	0	100	130	R(Pos.)
3	0	160	*	0	-110	0	R(Pos.)
4	-90	0	*	0	-90	90	R (Orient.)
5	0	0	*	72	0	0	R (Orient.)

IV. Assumptions

Our assumptions can be mentioned in two points;

1- We are interested in (arm) Manipulability, regardless orientation singularity.
2- We are interested in studying the non redundant Manipulator in x-y-z plane.

V. Simulation Strategy

In our research, we tried to test singular value of any serial manipulator according to these steps:
1- Finding the joint that has a vital role in changing the serial manipulator's configuration to singular one.
2- With simulation program; using robotic toolbox [7] making this joint ranging from one value to the other value.
3- Calculating the jacobian in every step in loop and Singular matrix.
4- Plotting every σ (normalized) and also the rank of the jacobian with number of configurations, robot has passed through changing from one singular phase to another. See fig (1)

$$\text{Normalized}\,\sigma_i = \frac{\sigma_i}{Max\{\sigma_{i1},\sigma_{i2},\sigma_{i3},.....,\sigma_{in}\}} \qquad (7)$$

i: Order of σ in the singular Matrix.
n: number of loops/ configurations during simulation.

5-Checking the rank of Jacobian while the serial robot in any other different configurations.

A Experimental Results for Puma560

The initial position is $Q_{initial}$ = [0, 0,-90, 0, 0, 0] degree and the final position is Q_{final} = [0, 0, 90, 0, 0, 0] degree.

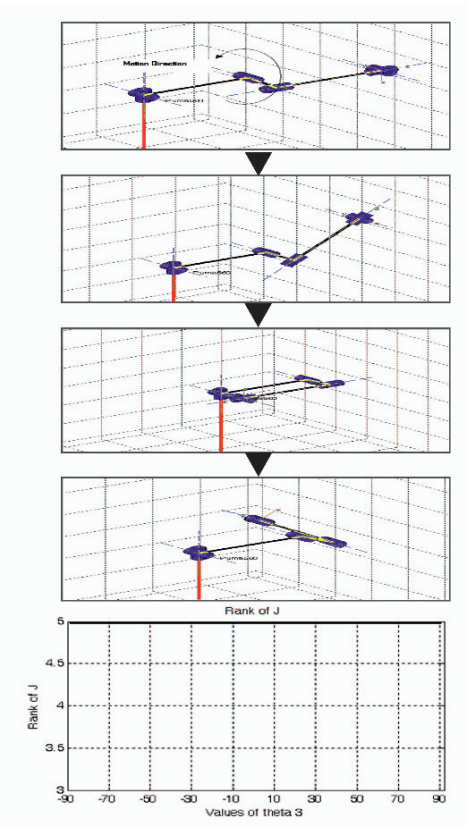


Fig1. Phases of Puma 560 in our experiment changing from one singular configuration to the other.

In Fig.2-a and fig 2-b, It is obvious that σ5 is exactly expressing the Manipulability's index. Also, rank of the jacobian during the experiment was constant at 5 (joint 6 and joint 4 were on same axis during the experiment).
In Fig.3, We have used another algorithm, depending on forward kinematic to generate the whole workspace of any manipulator in 2-D. Algorithm is:
1- Forward Kinematics
2- With every point, calculate Jacobian matrix.
3- Calculate SVD
4- Record in a matrix [Qs x y z σ_K]; K=min (rank (J)) [in all singular configuration, assuming every joint has a range from 0 to 360 Degree].
5- Plotting the coordinates with different colors according to σ_5 see fig (3).

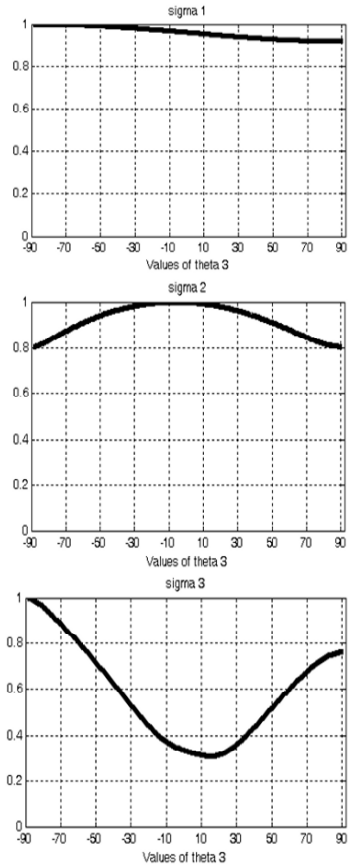


Fig. 2-a. The Behavior of **σ** 1 to **σ** 3 during the experiment

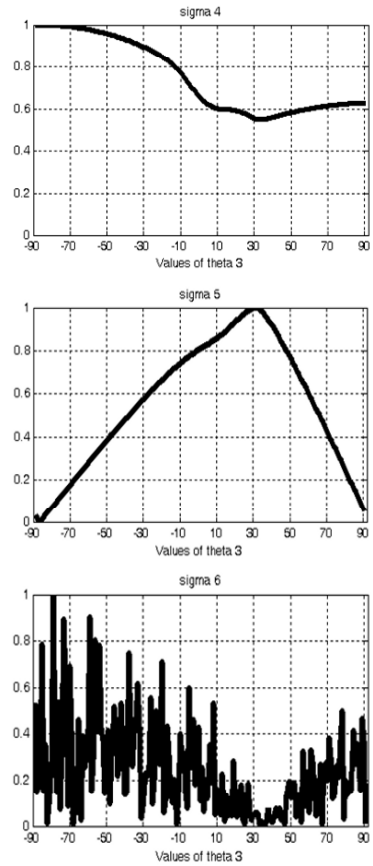


Fig. 2-b. The Behavior of **σ** 4 to σ 6 during the experiment

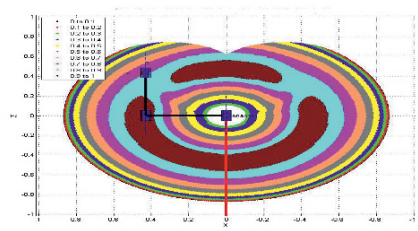


Fig.3. Manipulability's Bands of Puma 560 in 2-D workspace according to $\boldsymbol{\sigma_5}$

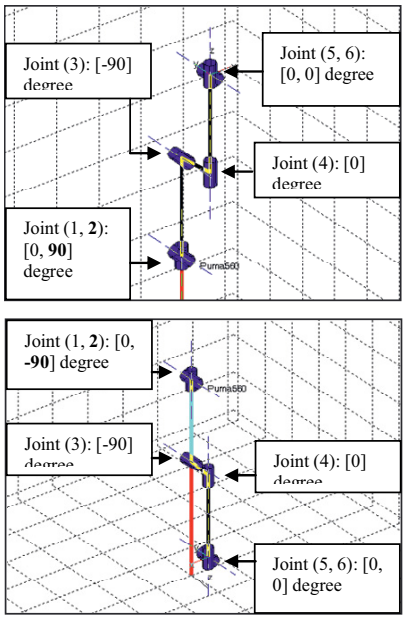


Fig.4. Two other Singular Configurations of Puma560 manipulator

Similarly, we did same strategy for Misubishi Movemaster® and Regular six degrees of freedom manipulator.

B Experimental Results of Six degrees of freedom Serial Manipulator

The initial position is $Q_{initial}$ = [0, 0,-90, 0, 0, 0] degree and the final position is Q_{final} = [0, 0, 90, 0, 0, 0] degree.

It is obvious from fig.6 that σ_3 is the only singular value in singular matrix that expressing exactly the arm Manipulability (fig.7 is supporting that by showing Manipulability bands perfectly).We have also observed the rank of the jacobian matrix during these phases was five when the manipulator was not in singular value and was four when the manipulator in singular value. Also, we have checked the rank of jacobian in other singular configuration as in fig.8. It was a surprise that the rank of the jacobian matrix has been reduced to three.

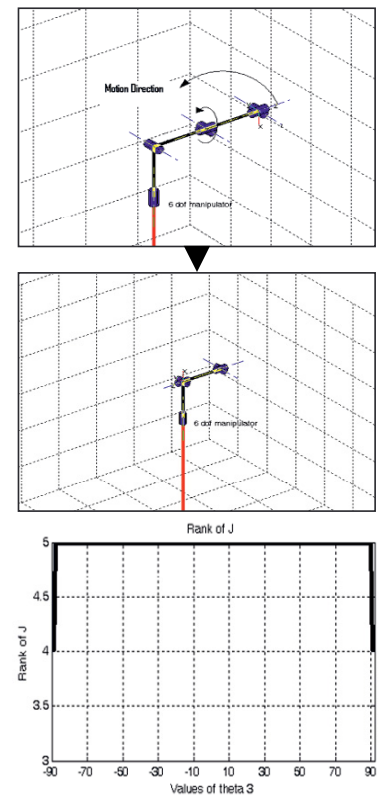


Fig.5. Phases of sex degrees of freedom manipulator in our experiment changing from one singular configuration to the other.

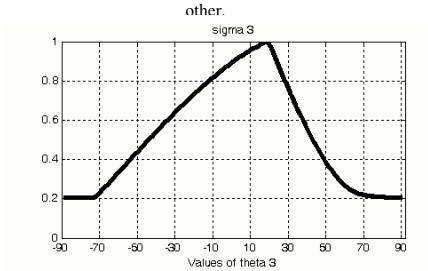


Fig.6. the Behavior of σ_3 during the experiment

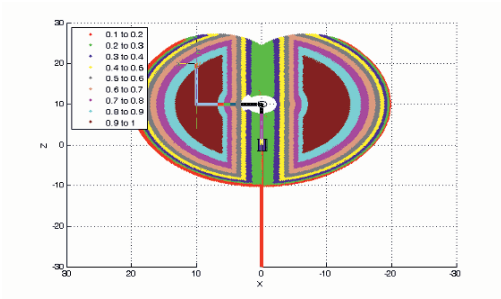


Fig.7. Manipulability's Bands of six degrees of freedom manipulator 2-D

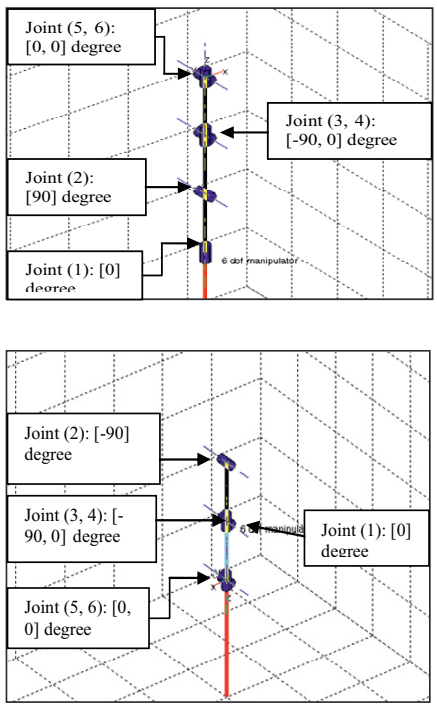


Fig.8. Two other Singular Configurations of six degrees of freedom manipulator

C Experimental Results Mitsubishi Movemaster ®

The initial position is $Q_{initial}$ = [0, 0, 0, 90, 0] degree and the final position is Q_{final} = [0, 0, 180, 90, 0] degree.

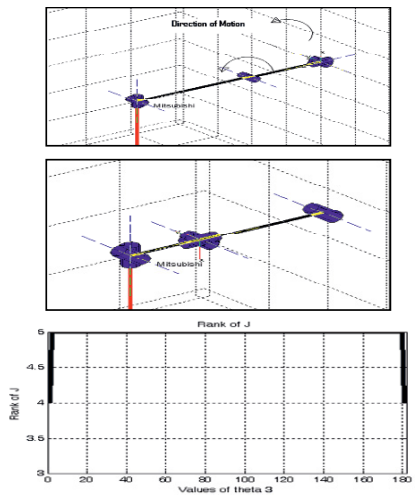


Fig.9. Phases of Mitsubishi Movemaster® in our experiment changing from one singular configuration to the other.

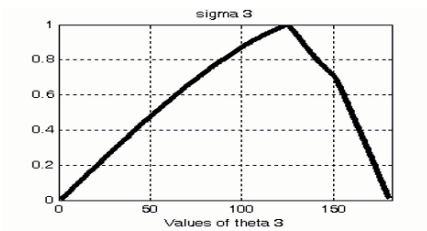


Fig.10. the Behavior of σ3 during the experiment

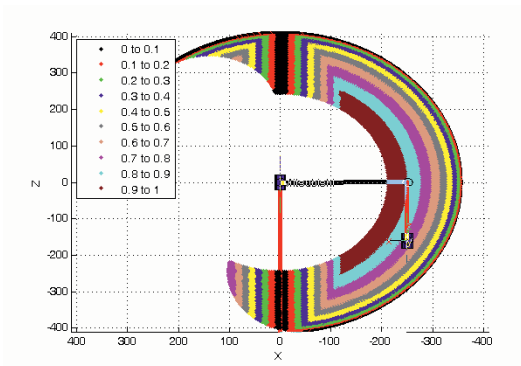


Fig: 11. Manipulability's Bands of Mitsubishi movemaster in 2-D workspace.

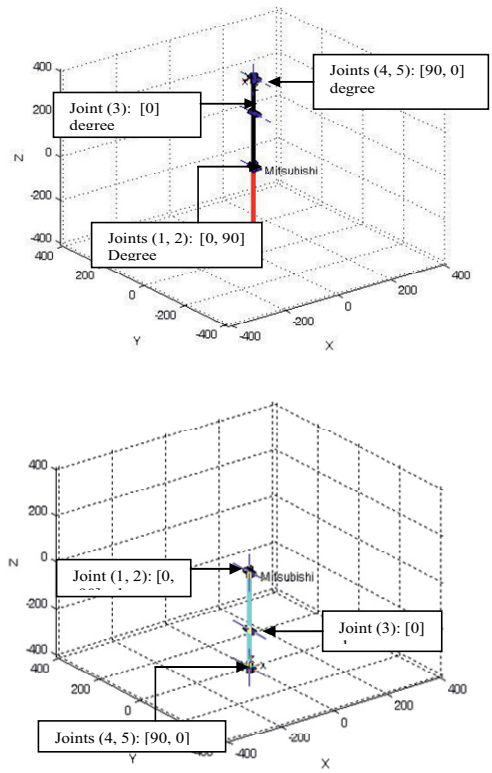


Fig.12. Two other Singular Configurations of Mitsubishi manipulator

TABLE 4
Summary of Results

Types Of Manipulator has been used in the Experiment	Order of **σ** that is expressing the Manipulability	Rank during changing the configuration from one singular to another	The least rank can robot reach in any other singular configurations
Puma560®	5	5	5
Six degrees of freedom	3	5 and 4	3
Mitsubishi Movemaster®	3	5 and 4	3

IX. Conclusion

It is obvious from the table 4 that the order of σ that expresses kinematic manipulability 's index or bands equals to the minimum rank of the jacobian matrix, that robot can reach in any singular Configurations , regardless the joints' limits.

X. Future Work

1- We will discuss in the future how to prove this relation mathematically.
2- As we mentioned above we have to optimize all points with deferent Manipulability's index to find max value of one of the σs and normalize to make the index's values ranges from zero to one. We will try to optimize the jacobian symbolically, and optimize the singular value decomposition also symbolically to find the differentiation of this value to optimize the maximum value.
3- What is the best location to fix a new robot's base already has a task to cover all points on trajectory with high Manipulability's index for every point. For instance, it was in past big argument about what is the best location to fix a manipulator will be used for painting the F-116 fighting plane.
4- Actually, we are thinking about new concept, we have not called yet, about resultant Manipulability between two serial manipulators, interacting with each other to do one task. For instance, one is holding the work piece and the other is doing some machining or welding operations.
5- Same concept can be taken into consideration ,when two manipulators are fixed on two different mobile robots or prismatic joint, holding a spherical or difficult symmetric object for instance, ball .From Kinematic Manipulability , what is the best posture with high Manipulability far enough from singular configuration to hold this object?
6- Now, kinematic manipulability becomes mature enough to be involved in any commercial software for robot simulation such as: Delmmia ®, Easy Rob ®..Etc

References

[1] P. H. Chang, "A Dexterity Measure for the kinematic Control of Robot Manipulators with Redundancy", February, 1988.
[2] I-Min Chen, "theory and application of modular reconfigurable robotic systems", 1994.
[3] T. Tanev, and I. Chandrov, " Performance evaluation of manipulation system and graphical representation of the characteristics", 1998.
[4] T. Tanev, and B. Stoyanov, "On the Performance Indexes for Robot Manipulators", 2000.
[5] H. Bruyninckx," The Robotics WEBook" , 2006.
[6] M. W. Spong , S. Hutchinson , and M. Vidyasagar, "Robot modeling and control", 2005.
[7] P. I.Corke, "Robotic toolbox", April 2002.

Region of Interest Labeling of Ultrasound Abdominal Images Using Hausdorff Distance

Naveen Aggarwal*,
CSE Deptt., Univ. Instt. Of Engg. & Technology, Panjab University,
navagg@gmail.com

Prof. Nupur Prakash,
School of Information Technology, Guru Gobind Singh Indraprastha University, Delhi

Prof. Sanjeev Sofat,
CSE Dept., Punjab Engineering. College, Chandigarh

***Abstract*- This paper presents a two stage approach to segment the ultrasound abdominal images. During first stage, images are segmented into different regions which are stored in the database along with their statistical properties. In second stage, different segmented regions in the database are used for Region of Interest labeling of ultrasound images using hausdorff distance. The quality of ultrasound images are strongly affected by the presence of speckle noise. As Speckle noise is multiplicative in nature, so homomorphic filtering is found to be best suited to reduce such a noise from the images. To segment out the region for the learning purpose, improved marker-controlled watershed segmentation algorithm is used. But this normally results in over segmentation. This over segmentation problem is solved by defining the internal markers on the image using morphological operations. All the regions segmented out in this step are stored in the database along with their statistical properties. Finally, a template matching approach is proposed to match the regions of a given ultrasound image with already stored template regions using hausdorff distance. If a region in the image matches with the template, it is appropriately labeled as per the other details stored in the database. It is observed that the efficiency of whole system depends upon the efficiency of learning stage. Although the computational complexity of system is very high during learning stage, but it is relatively very less during implementation stage. The overall efficiency of the system is found to be very good in terms of precision and recall parameters.**

I. INTRODUCTION

The process of dividing the image into set of non-overlapping regions is called segmentation. The semantics of these regions are defined by the criteria used to separate them. The success or failure of any technique depends upon the segmentation accuracy. The precision to which images are segmented also depends upon the problem being solved. Segmentation should stop when the object of interest are segmented out from the image. Most of the medical image analysis method requires segmentation as a preliminary step with a user input at the beginning to initialize the process. These programs do not require any user interaction afterwards. But the accuracy of these methods can always be improved with more user interaction. The uncertainty about the region of interest in medical images makes them unsuitable for completely automated process. None of the medical image segmentation method is completely automated [1] and the amount of user interaction defines the accuracy of the method.

Among Medical Images, Ultrasound image segmentations is an important problem due to the presence of strong speckle noise which makes it difficult for medical practitioner to recognize small cysts or lesions with a naked eye[2]. The multiplicative speckle noise cause so much of information loss that the accuracy of any ultrasound segmentation technique also depends upon the knowledge base behind the technique. This knowledge base can be generated by designing a learning based system or by increasing interaction with the medical experts. Hence a single technique is not a sufficient for segmenting all the different types of ultrasound images.

In this paper, we have proposed a technique for region of interest labeling and region detection technique in abdominal ultrasound images. An abdominal ultrasound produces a picture of the organs and other structures in the upper abdomen such as liver, gall bladder, kidney etc. The various deformities, tumors or stones in these organs can be detected with the help of the ultrasound images. In our system, we improve the quality of images by minimizing the speckle noise and a stepwise procedure is followed to segment out the region of interest. The proposed system is divided into two stages. The first stage is the learning stage in which a supervised database of segmented regions is created with the help of an expert. In the second stage, templates from the database are matched with the given ultrasound images using hausdorff distance method and the matched regions are appropriately labeled.

The rest of the paper is organized as follows. The related techniques have been discussed in the section II. Section III describes the proposed method. The analysis and results of proposed method are explained in the section IV. Finally the concluding remarks and future enhancements are discussed in section V.

II. RELATED TECHNIQUES

Different segmentation approaches used for medical imaging analysis can be classified in two main classes. First, on the basis of the amount of user interaction required by the particular approach and second, on the basis of the particular

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 192–197.

technique used to segment the image into different regions of interest.

In our first classification, different methods are recognized in two general categories, i.e. Rule based methods and Pattern recognition based methods. In Rule based methods, automated process performs on the set of predefined conditions or rules. Leemput et al. [3] has developed a rule based methods to classify pixel based on set of previous results where a digital model atlas of brain is generated. But generating such a model atlas is very cumbersome and complex task. Further a small error will reduce the efficiency of the whole system. Other similar approaches such as Kaus [4], Atkins [5] etc. [6, 7] uses statistical properties of pixels to classify them by taking advantages from the previous experiments. The second category of approaches in this class use predictive models and artificial learning techniques. In this, Marroquin et al [8] performed Bayesian analysis on the data for automatic registration of pixels to define intensity models for segmentation of brain tumors. The detail classification of data mining based approaches for medical image segmentation is provided by the Antonie et al [9].

In our second classification, most of the algorithms focus on main six approaches such as amplitude thresholding [10], component labeling [11], Boundary-based approaches [12], Region-based approaches and clustering [13, 14], Template matching [15] and Texture segmentation[16]. The amplitude thresholding based techniques segment the image into different regions based on the intensity values of pixels. But the presence of noise such as speckle noise in ultrasound images, greatly affect the efficiency of this approach. Hua et al [11] used the feature based component labeling approach after segmenting the image using thresholding techniques. The boundary based approaches that uses simple edge detection and linking algorithm are not efficient for ultrasound images. For this boundary points can be detected using continuous wavelet transformation [12] and then a curve around these points is obtained by using B-Splines [15]. But the resolution of ultrasound images is usually very less so recall and precision of boundary based approaches is very less. Region based and clustering approaches are better suited for low resolution and noisy images. Most of the region based approaches use the statistical properties of the image at global or local level to distinguish between regions [13, 14]. The statistical properties of images can also be used to define the texture of the image. But calculation of statistical value is very computational complex. Further the efficiency of these techniques depends upon the region merging and splitting criteria used for the segmentation.

Our proposed approach uniquely combines the features of both of the classes discussed above. A little user interaction in ROI labeling enhances the efficiency of overall system.

III. PROPOSED METHODOLOGY

The proposed methodology is illustrated in the Figure 1. It consists of five main stages. The first three of these stages, namely, Minimizing speckle noise, marker-controlled watershed segmentation, Region splitting and merging constitutes the learning part of the system and ends up in creation of the supervised database for the system. The fourth and fifth stage, namely, Hausdorff distance based matching and ROI labeling constitutes the implementation part of the system. Each of these stages is discussed in subsequent sections.

A. Minimizing Speckle Noise

Speckle noise can be considered as an infinite sum of independent, identical phasors with random amplitude and phase. This makes its amplitude a complex number and intensity field is magnitude of that complex number [15]. But the intensity field has exponential distribution with variance double than the actual amplitude. The speckle noise in the ultrasound images can be minimized by taking either spatial averaging of speckle or by using homomorphic filtering. As speckle noise is not additive, but multiplicative in nature, hence homomorphic filtering suits better as compared to spatial averaging.

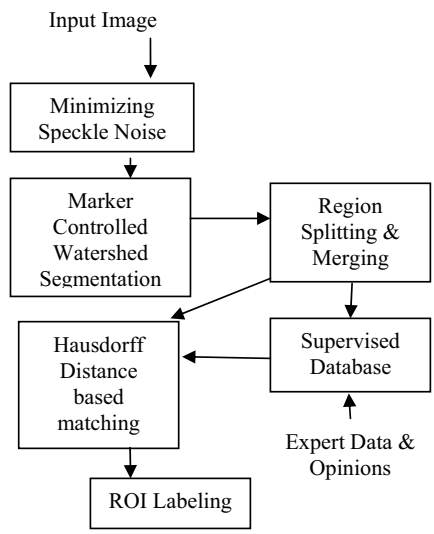


Figure 1: Block Diagram representing the different modules of the system.

In homomorphic filtering, we take the log of actual image, which results in separation of white noise component.

$$\text{Log (Original Image)} = \text{Log (Actual Image)} + \text{Log (White Noise)} \quad (1)$$

Now, a particular filter can be designed to remove the white noise component and then exponential is taken to get actual image back. As suggested in [15], homomorphic wiener filter performs quite well as compared to the linear filtering or other homomorphic filters. Figure 2. describes the actual process used.

B. Marker Controlled Watershed Segmentation

The watershed segmentation is applied to recognize the different segments of the image based on their topographical details. As the distance transforms are not easily applicable to the ultrasound images, so we have used the gradients for watershed segmentation. To avoid the over segmentation of the image, morphological operators are used to mark the desired regions in the image.

For each of the segments recognized above, its statistical properties such as mean and variance are calculated. Now, for the ensemble of ultrasound images taken for a particular patient, each region is represented by the correlation coefficient and covariance values.

C. Region Splitting & Merging

Above calculated correlation coefficients and covariance values are taken as predicate conditions for splitting the image into regions. Different steps used for region splitting and merging are as follows.

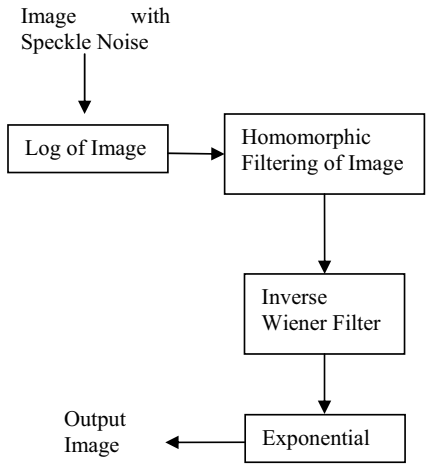


Figure 2: Process of Homomorphic filtering

1. Let R represents the entire image region and predicate P represents the set of all predicate conditions.
2. For each of the predicate conditions, it is checked whether, P(R) is true or false.
3. If it is false, then whole region is again divided into four quadrants and then for each quadrant same step if performed.
4. Finally, adjacent regions are merged together, if they satisfy the predicate conditions.

Efficiency of such a procedure depends upon the correctness of predicate conditions. As it is assumed that the predicate conditions are formed from the statistical properties of ensemble of watershed segmented ultrasound images. Hence overall efficiency is very good as compared to other segmentation algorithms. Further rather than storing the all the ultrasound images corresponding to a particular patient, only the segmented regions along with their statistical properties are stored in the database. But before storing the data, experts can interact with the system to label the region as per its properties. The expert can reject the wrongly segmented region and add only the region having accuracy more than preset threshold value.

D. Hausdorff Distance based Matching

The Hausdorff distance is a metric between two set of points. Let the two-dimensional point sets U and V denote representation of image area of interest and object template respectively. Each point of the set stands for certain feature in the image and an edge point is taken in our case. Now the hausdorff distance is defined as

H(U,V)= max(h(U,V), h(V,U)), where

$$h_{\mathrm{mod}}(U,V) = \frac{1}{|U|}\sum_{u\in U}\min_{v\in V}\|u - v\| \quad \text{as}$$

suggested by Dubuisson[20] to take the average of single point distances to reduce the impact of outliers for making it suitable in our domain of ultrasound images. The goal is to find the transformation parameters such that the hausdorff distance between the transformed model Tp(V) and U is minimized. The parameter space is limited to two transformations only i.e. translation and scaling. So the detection problem can be described as

$$d_p = \min_{p\in P} H(U, Tp(V))$$

The regions detected are then appropriately labeled as per the properties of that region stored already in the database. Only the segmented image region and stored object templates are matched. The parameter space is also limited around the segments to further reduce the complexity.

IV. RESULTS

The proposed methodology is tested on more than 80 ultrasound images. The different algorithms are implemented in C and called from Matlab environment. Final results are evaluated in Matlab 7.0. Figure 3 illustrates the application of homomorphic filter on ultrasound image affected with speckle noise. Its results are also compared with the adaptive median filtered [15] image. From the results, it is quite clear that; homomorphic filter works better because of presence of multiplicative rather additive speckle noise. The adaptive median filter can be used for additive noise only. Now images are corrected from any degradation by applying the inverse wiener filter. This step will further enhance the sharpness of images. The result of this is shown in Fig 4(a). Before applying the watershed segmentation, gradient of the image is taken which is shown in Figure 4(b). The gradient of image is taken by applying the modified sobel filter to the image. Modification in sobel filter is done by giving more weight age to the centre pixels as compared to the others. Now the watershed segmentation is applied on this gradient image rather than on the actual image. The over segmentation of image due to watershed segmentation is quite visible in the Figure 4 (c). For obtaining Figure 4 (c), normal watershed

segmented image is complemented and then superimposed onto the original image to compare the results. Figure 5 describes the application of marker- controlled watershed segmentation of images. The efficiency of this step affects the efficiency of whole system. The better the marker image used, better is the output of segmentation. Different marker images based on the morphological operations are created and used in the system. It is observed that no single marker image is suitable for the different types of ultrasound images.

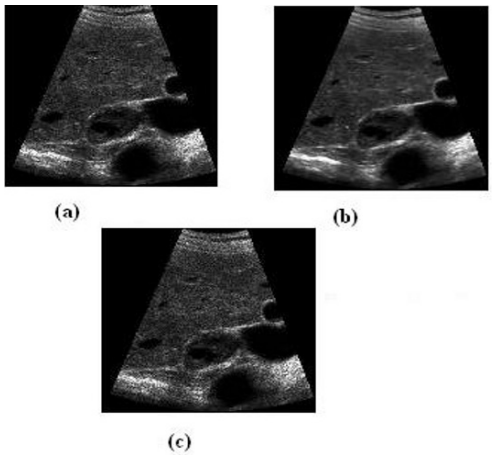


Figure 3: Comparison of Homomorphic filtering with the Adaptive median Filtering. (a) Represents the image corrupted with speckle noise (b) Result of Homomorphic filtering (c) Result of Adaptive Median Filtering

In Figure 5, all the different types of marker images used in the system are displayed. Hence a small user interaction at this stage can remarkably improve the efficiency of the segmentation step. In Figure 5(a) marker image is created by performing dilation operation on it. The dilation process grows and thickens the objects in the image as defined by the structuring element. The properties of this structuring element can be adjusted as per the abdominal ultrasound images.

In Figure 5(b), marker image is created by performing the erosion operation on the background pixels. The erosion operation shrinks and thins the boundary values in the image. A different marker image can be obtained, if the foreground objects are marked by opening the image. Now this opened image is superimposed on the actual image to get final marked image which is shown in Fig 5 (c). Similarly in Figure 5(d), both foreground and background objects are marked by successive opening and closing of the images. Both of these are superimposed on the original image to get the marked image. All the above methods discussed above to create the marker image are very computational expensive.

To reduce the computational complexity of the whole system, simple gradient can also be used as marker image. Results of segmentation are further improved by using the region splitting and merging techniques. The final segmented regions are stored in the database by neglecting small regions of less than 5 pixel size. Finally in Figure 6, displays the ROI labeling results after matching the object templates from already stored database using hausdorff distance.

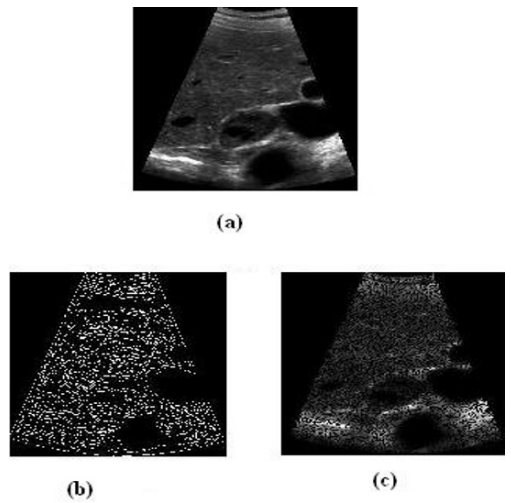


Figure 4: (a) Improved Image after applying inverse wiener filter (b) Gradient of original image (c) Over segmentation caused by watershed algorithm

This database is again prepared with the help of medical expert. Once the region is detected, its characteristics are displayed along with the region. The accuracy of system can be checked either subjectively from the expert opinions or objectively by using two parameters,

$$\text{Re}\,call = \frac{Total\_Correct\_Detections}{Total\_Existing\_Events} \quad \text{and}$$

$$\text{Pr}\,ecision = \frac{Total\_Correct\_Detections}{Total\_Detections}$$

If there is an overlap of at least 0.5 with the labeled region to the region labeled by the expert, then the system is assumed to be correct. Otherwise, it is considered as wrong detection. It is found that most of the results are overlap of more than 0.75. Precision and Recall is calculated for all the images for different regions as mentioned by the medical expert. These results are summarized in Table I.

All the results shown here are based on very small initial data. The reason of less efficiency is due to the insufficient learning of the system during initial stages of execution. As the system grows with more learning through more data, its efficiency is also enhanced.

V. CONCLUSION

A method to segment the ultrasound abdominal images and ROI labeling is proposed. This method goes through two stages. In first stage, speckle noise from the images is removed and marker controlled watershed segmentation is applied.

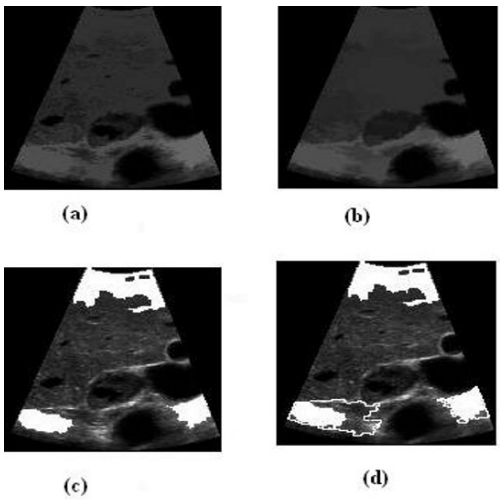


Figure 5: (a) Marker Image obtained by performing Dilation (b) Marker Image obtained by performing Erosion (c) Foreground superimposed Marker Image (d) Background and Foreground both are superimposed on Original image

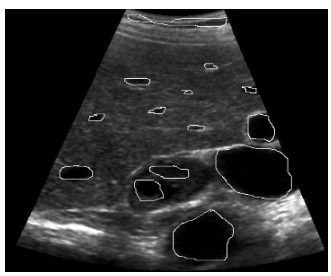


Figure 6: ROI Labeled Image as per the template stored in the region.

TABLE I

SUMMARY OF PRECISION AND RECALL VALUES FOR DIFFERENT REGIONS

Region	Recall	Precision
Region 1	0.88	0.91
Region2	0.92	0.86
Region3	0.84	0.75
Region 4	0.93	0.90

In this stage, finally regions are segmented based on their statistical properties and a database is created. The different segmented regions in this database are used as templates for ROI labeling in the final stage. As the data grows, with more learning of data, the computational efficiency of final stage is greatly enhanced. The learning stage needs to be improved by using advance learning methodologies for large amount of data.

REFERENCES

[1] J. L. Foo. A survey of user interaction and automation in medical image segmentation methods. Technical Report ISU-HCI-2006-2, Iowa State Universtity - Human Computer Interaction, 2006.

[2] Xiaohui Hao, Charles Bruce, Cristina Pislaru and James, F. Greenleaf, "A Novel Region Growing Method for Segmenting Ultrasound Images," IEEE Ultrasonics Symposium, vol. 2, pp. 1717-1720, 2000.

[3] Leemput K.V.; Maes F.; Vandermeulen D.; Suetens P. Automated Model-Based Tissue Classification of MR Images of the Brain. IEEE Transactions on Medical Imaging 1999 Oct; 18(10):897-908.

[4] Kaus M.R.; Warfield S.K.; Navabi A.; Black P.M.; Jolensz F.A.; Kikinis R. Automated Segmentation of MR Images of Brain Tumors. Radiology 2001;

[5] Atkins M.S.; Mackiewich B.T. Fully Automatic Segmentation of the Brain in MRI. IEEE Transactions on Medical Imaging 1998 Feb; 17(1).

[6] Hartmann S.L., Parks M.H., Martin P.R.; Dawant B.M., Automatic 3-D Segmentation of Internal Structures of Head in MRI using a Combination of Similarity & Free-Form Transformations, Part II, Validation on Severely Atrophied Brains. IEEE Trans. on Medical Imaging 1999.

[7] Soler L.; Delingette H.; Malandain G.; Montagnat J.; Ayache N.; Koehl C.; Dourthe O.; Malassagne B.; Smith M.; Mutter D.; Marescaux J. Fully Automatic Anatomical, Pathological, and Functional Segmentation from CT Scans for Hepatic Surgery. Proceedings of SPIE Medical Imaging Symposium; February 2000; San Diego, CA.

[8] Marroquin J.L.; Vemuri B.C.; Botello S.; Calderon F.; Fernandez-Bouzas A. An Accurate and Efficient Bayesian Method for Automatic Segmentation of Brain MRI. IEEE Transactions on Medical Imaging 2002.

[9] Antonie M.; Zaiane O.R.; Coman A. Application of Data Mining Techniques for Medical Image Classification. Proceedings of the Second International Workshop on Multimedia Data Mining (MDM/KDD'2001) with ACM SIGKDD 2001.

[10] R.C.Gonzalez, R.E.Woods, S.L.Eddins, Digital Image Processing usig Matlab, Pearson, 2005.

[11] Hiransakolwong, N., Hua, K.A., Khanh Vu, Windyga, P.S., Segmentation of ultrasound liver images: an automatic approach , in proceeding of International conference on Multimedia and Expo (ICME), Pages: 573-6, 2003.

[12] Yoshida, H., Keserci, B., Casalino, D.D., Coskun, A., Ozturk, O., Savranlar, A., Segmentation of liver tumors in ultrasound images based on scale-space analysis of the continuous wavelet transform, in proceedings of IEEE Ultrasonics Symposium, Pages:1713 - 1716 , 1998.

[13] Xiaohui Hao, Charles Bruce, Cristina Pislaru and James F. Greenleaf, "A Novel Region Growing Method for Segmenting Ultrasound Images," IEEE Ultrasonics Symposium, vol. 2, pp. 1717-1720, 2000.

[14] Ashish Thakur, Radhey Shyam Anand, A Local statistics based region growing segmentation method for ultrasound medical images, in International journal of signal processing, vol 1, number2, pages 141-146, Enformatika, 2004.

[15] Anil K. Jain, Fundamentals of Digital Image Processing, Pearson Education, 2005.

[16] Rong Lu, Yi Shen, Automatic Ultrasound Image Segmentation by Active Contour Model Based on Texture in the proceedings of First International Conference on Innovative Computing, Information and Control - Volume II (ICICIC'06) pages 689-692, 2006.

[17] Wang,W., Song, Y., Zhang.A "Semantics-based image retrieval by region saliency", in International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, Springer 2002.

[18] Mihai Datcu, Klaus Seidel, "An Innovative Concept For Image Information Mining", in proceedings of Int'l workshop on Multimedia Data Mining with ACM SIGKDD, 2002.

[19] Osmar R. Zaiane, Jiawei Han, Ze-Nian Li, Jean Hou, "Mining Multimedia Data", MDM/KDD2001, Second International Workshop on Multimedia Data Mining in conjunction with Seventh ACM SIGKDD International Conference on Knowledge Discovery & Data Mining , San Francisco, 2001.

[20] M.P.Dubuisson and A.K.Jain, " A modified Hausdorf distance for object matching" in ICPR94, pages A:566-568, Jerusalem, Israel, 1994.

Control of Electric Motor Parameters on the Basis of QR- Decomposition Technique

First A. Viktor Melnikov, Second B. Yelena Zigangirova, and Third C. Nikolay Shishkov,

***Abstract*—Control of DC motor on the basis of QR-decomposition technique is given in this paper. Authors propose to control DC motor parameter using step-by-step method of rotor's parameters definition. The first step is definition of main poles magnetic flux and active resistance of exciting circuit. Active resistance, rotary rotations are defined on the second step. Appropriate equations of transforming are given in the article. Identification results using this technique are presented on the graphs. Also the results of simulation are shown in the paper.**

I. Introduction

THE basis of control is computer processing of information which consist in voltage and current of electric motor. The simulator of electric motor is needed for computer processing of information [1].

Control is necessary because the majority of electric motor's electromagnetic parameters and state variables are inaccessible for direct measurements during the process of function. For instance, such parameters for DC motor are active resistance and inductance of rotor and field coil. Some parameters such as magnetic flux in air gap of DC motor, electromagnetic moment and rotary rotations can be measured but it is preferable calculate due to difficult operating mode of electric drive.

Moreover, DC motor's parameters depend on operating mode and it's heat state. So, active resistance may change more than 1.4 times and inductance changes on 30-40% in direct start of operating mode. Active resistance depends on DC motor's heat state and can change during the its operating mode. It is typical for recursive short-time mode.

It appears from this that current data of electric motor's parameters are necessary to define during the operating mode of electric drive. It is possible by using control of parameters and indirect control of electric motor's coordinates included in definition of current electromagnetic data and variable data of electric motor velocity in real-time during the operating mode.

F. A. Author is with the Pavlodar Innovational Eurasian University, The Energy department, 140000 Pavlodar, Kazakhstan, Lomov str, Office 45, tel. 8(3182)452779 (e-mail: s-melnik@yandex.ru).

S. B. Author is with the Pavlodar Innovational Eurasian University, The Energy department, 140000 Pavlodar, Kazakhstan, Lomov str, Office 45, tel. 8(3182)452779 (e-mail: lena_khaled@hotmail.com).

T. C. Author is with the Pavlodar Innovational Eurasian University, The Energy department,140000 Pavlodar, Kazakhstan, Lomov str, Office 45, tel. 8(3182)452779 (e-mail: serg_nikoni@rambler.ru).

II. Main

Typically universal model on the basis of generalized electric machine (GEM) are used to analyze dynamics of electromechanical conversion of energy. The universal models consist of fundamental electromagnetic field theory. Using GEM it can be defined simulator of any inductive electric machine.

Using equation of electric drive motion [1] it can be formed following equations:

$$\begin{cases} -U_a = -e + (R_a + R_f)\cdot i_a + L_f \cdot \dfrac{di_a}{dt} + W_f \dfrac{d\Phi}{dt}; \\ W_f \dfrac{d\Phi}{dt} = L_f \cdot \dfrac{di_a}{dt}; \\ M_e = C_м \cdot i_a \cdot \Phi; \\ e = C_e \cdot n \cdot \Phi = \dfrac{L_f \cdot i_a \cdot n \cdot C_e}{W_f}; \\ J \dfrac{dn}{dt} = M_e - M_{st}; \\ \dfrac{i_a^2 \cdot (R_a + R_f)}{Am} = \theta_м + T_\theta \cdot \dfrac{d\theta}{dt}. \end{cases} \quad (1)$$

U_a - instantaneous voltage, V;

$å$ - EMF of rotation, V;

R_a - rotor active resistance, Ohm

R_f - active resistance of exciting circuit, Ohm;

i_a - current , A;

$i_{a.k.}$ - current component, A;

W_f - number of turns of exciting circuit;

L_a - inductance of rotor, H;

L_f - inductance of exciting circuit, H;

M_e - electromagnetic moment, *N*m*;

$ì_{st}$ - static moment, *N*m*;

J - moment of inertia, $kg*m^2$;

$C_e, C_м$ - coefficients depend on electric motor [1];

$\hat{O}$ - magnetic flux, Wb;

Am - heat emission coefficient of electric motor;

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 198–200.

T_θ - heat constant of electric motor;

$\theta_м$ - electric motor temperature, ^0C;

Active resistance and inductance of exciting circuit, active resistance and inductance of rotor of DC motor are calculated using the QR-decomposition technique [2, 3] by two steps. On the first step equations (1) and measured current and field coil voltage and first derivative current are formed into arrays of current, it's first derivative and field coil voltage. On the second step measured current, its first derivative and rotor voltage are formed into array these data.

For derivation of estimate on the basis of QR-decomposition technique equation (1) is transformed in to following:

$$-U_f = a_1 \cdot i_a + a_2 \cdot \frac{di_a}{dt}; \quad (2)$$

Where $a_1=R_{f;}$ $a_2=L_{f;}$

Than equation (2) transforms to matrix form of identifiable object:

$$Ax=b; \quad (3)$$

Where

$A=\left[|i_a|;\left|\frac{di_a}{dt}\right|\right]; b=\left[-U_f\right]$ - vectors of observable output and input variables;

$x=\begin{bmatrix} r_f \\ L_{f.} \end{bmatrix}$ - vector of assessed parameters.

Evaluation of parameters vector are made by solution of equation:

$$[QR]=qr([A]); \quad (4)$$

$$\begin{bmatrix} r_f \\ L_f \end{bmatrix}=|R|\backslash(|Q|^T \times |b|), \quad (5)$$

Where r_f, L_f - active resistance and inductance of exciting circuit,

$|R|$ – upper triangular matrix,

$|Q|^{\mathrm{T}}$ – orthogonal matrix,

$|b|$ – array of voltage;

Matrix $|Q|$ and $|R|$ are defined by QR-decomposition technique:

$$[|Q|,|R|]=qr\left(\left[|i_a|,\left|\frac{di_a}{dt}\right|\right]\right), \quad (6)$$

The value of magnetic flux of main poles is defined by using the obtained inductance L_f

$$\hat{O}=\frac{L_f \cdot i_a}{W_f \cdot p},$$

Where p – number of pole pair of electric motor;

The results of simulation are shown on figure 1 and 2.

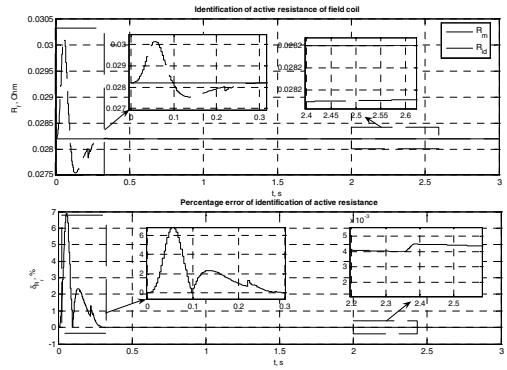


Fig. 1. Control results of active resistance of field coil; ___ model, --- identified

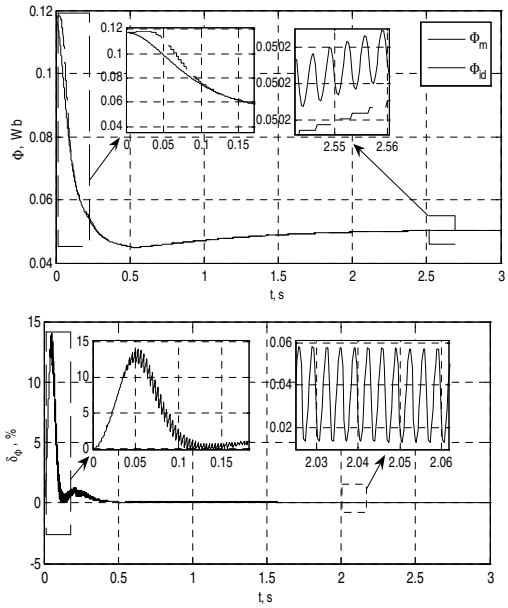


Fig. 2. Control results of magnetic flux of main poles; ___ model, --- identified

On the second step for derivation of estimation on the basis of QR-decomposition technique equation (1) is transformed into voltage equation:

$$-U_a - R_a \cdot i_a = a_3 \cdot \frac{di_a}{dt} - a_4 \cdot \left(\frac{i_{a.k.} \cdot L_f \cdot \tilde{N}_a}{W_f}\right), \quad (7)$$

Where: $a_3=L_{f;}$ $a_4=n_{;}$

R_a - Parameter, defined on processing mode with static rotor [1]

Than equation (7) transforms to matrix form of identifiable object:

$$A_1 x_1 = b_1, \tag{8}$$

Where $A_1 = \left[\left|\frac{di_a}{dt}\right|; \left|\frac{i_{a.k.} \cdot L_f \cdot \tilde{N}_{\dot{a}}}{W_f}\right|\right]; b_1 = \left[-U_a - R_a \cdot |i_a|\right]$ - vectors of observable output and input variables;

$x_1 = \begin{bmatrix} L_a \\ n \end{bmatrix}$ - vector of assessed parameters.

Evaluation of parameters vector are made by solution of equation:

$$[QR] = qr([A_1]); \tag{9}$$

$$\begin{bmatrix} L_a \\ n \end{bmatrix} = |R| \backslash (|Q|^T \times |b_1|), \tag{10}$$

Where L_f - inductance of electric motor rotor,

n - rotary rotations,

$|R|$ - upper triangular matrix,

$|Q|^T$ – orthogonal matrix,

$|b_1|$ - array of voltage;

Matrix $|Q|$ and $|R|$ are defined by QR-decomposition technique:

$$[|Q|,|R|] = qr\left(\left[\left|\frac{di_a}{dt}\right|; \left|\frac{i_{a.k.} \cdot L_f \cdot \tilde{N}_{\dot{a}}}{W_f}\right|\right]\right). \tag{11}$$

The results of simulation are shown on figure 3.

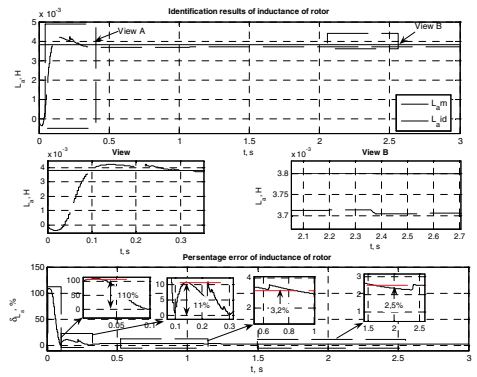


Fig. 3. Identification results of inductance of rotor;___ model, --- identified

III. CONCLUSION

According to the results we can say that QR-decomposition technique is adaptable for control of DC motor parameters in operating mode. In steady-state mode the control relative error is not more than 0,5 % for field coil and not more than 3% for inductance of rotor. Number of unknown parameters of equation 7 is increased due to difficulties of rotor characterization. It causes to singular matrix and inaccuracy. This problem can be solved by using step-by-step method of rotor's parameters definition. The first step is definition of rotor inductance and velocity. Active resistance of rotor parameter is known [1]. Assumed that inductance parameter is known (step 1), active resistance, rotary rotation are defined on the second step.

REFERENCES

[1] Melnikov, V, Kibartas, V, Shishkov, N. Identification of parameters of DC motor // Science and education in XXI century – 2006. – Vol. 3. – №1. – P. 112–113..

[2] Slock D.T.M., Kailath T. Numerically stable fast transversal filters for recursive least squares adaptive filtering // IEEE Trans. Signal Processing. – 1991. – Vol. 39. – №1. – P. 92–114.

[3] Papaodysseus C. A robust, parallelizable, O(m), a posteriori recursive least squares algorithm for efficient adaptive filtering // IEEE Trans. Signal Processing. – 1999. – Vol. 47. – №9. – P. 2552–2558

Navigation of Mobile Robots Using 6DOF SLAM Accelerated by Leveled Maps

Ondrej Jez
Dept. of Control and Instrumentation
Faculty of Electrical Engineering and Communication
Brno University of Technology
Kolejni 2906/4
Brno 612 00, Czech Republic
ondrej.jez@phd.feec.vutbr.cz

Abstract- **This article is presenting a method for simultaneous localization an mapping (SLAM) of mobile robots in unknown environments. The method is using 3D range data which are acquired from a continuously inclined laser time of flight ranging sensor. In order to accelerate existing computational time-demanding Iterative Closest Point SLAM methods, the SLAM procedure is using a leveled map of the environment created from the 3D data which serves as a pre-estimation of the optimal translation and rotation in three degrees of freedom. This can eliminate the number of iterations of the final robust six degrees of freedom (6DOF) ICP registration, thus saving a lot of computational time. The criteria of the success of this method are the characteristics of the environment, especially the way the environment is structured, and the appropriate setting of the leveled map extraction. Results from the testing of the method in indoor environment are also presented.**

I. Introduction

This article presents an ICP based method of simultaneous localization and mapping of mobile robots in unknown environments. The motivation of the development of a navigation method is that it is an essential step towards the autonomy of a robot. Only after the robot can localize itself in the environment and sense its surroundings in an organized way, it could perform various tasks which could range from utility of households, transportation of people and goods, exploration of unknown environments, automatic civil construction etc.

This research is focusing on the localization and map building using 3D range measurements. Since our method for localization and map building is to be operating in unknown environments, both tasks are inseparable and thus a Simultaneous Localization and Mapping – SLAM algorithm is being used, which in our case of three dimensional space is done in six degrees of freedom. Currently existing methods' main issue is the computational cost of the registration of different 3D range images of the environment; therefore the main focus of this research was to accelerate the existing methods. The method we are presenting here is a two stage ICP algorithm using a leveled map to pre-estimate the optimal translation and rotation in three degrees of freedom (3DOF), followed by a robust 6DOF registration.

In the following sections, the state of technology in 3D range data acquisition will be first described, followed by an insight into the theory of the leveled map accelerated 6DOF ICP SLAM algorithm and finally the so far achieved experimental results will be presented.

II. State of Technology in 3D Range Image Acquisition

Today, mobile robots are equipped with a variety of sensing devices such as cameras, range sensors (ultrasound, optical, radio), tactile sensors, gyros etc. which allow solving the task of localization and exploration. Some of these sensing devices are able to produce a 3D model of the environment, allowing the exploration of the surroundings in such a precision and detail that the possibilities of using this information to navigate the robot are so far unused.

A. Stereo Vision

This subsection will be commenced by discussing a minority approach to 3D mapping based on the use of stereo vision by CCD cameras. This approach is applied e.g. in research by Se et al. and Biber et al. [7][2]. Extracting 3D information from stereo vision means in fact estimating depth from two images. This is usually done using the triangulation method. The main constraint and also the main reason, why this method is quite rare in the robotics research, is that the stereo base line is limited by the robot's dimensions and often cannot exceed more than a half of meter. This together with relatively low resolution of currently available cameras leads to very high imprecision and range measurement error, often being about ten percent in distances up to ten meters (the error gradually grows with growing distance).

B. Use of Time of Flight 2D and 3D Laser Rangefinders

The current mainstream in 3D mapping is based on the use of 2D laser rangefinders, which are installed in an actuating mechanism allowing the inclination of the sensor and thereafter 3D range data acquisition. The main advantages of a 2D laser rangefinders is high precision of measurements over long distances (often 80m) while the minimal measured distance with current technologies is quite low (down to 2cm).

There are more ways of inclining the ranging sensors in order to obtain the 3D data - different team's solutions will

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 201–206.

now be discussed. The most comprehensive analysis of the application of 2D laser rangefinder for fast 3D scanning by rotating it around different axes was performed by Oliver Wulf from the University of Hannover [13]. The team from the ISE/RTS department has tested four methods of inclining the SICK laser rangefinder: pitching, rolling, and two different yawing methods. The analysis focuses mainly on the measurement of point distribution of various inclination principles, while the density of measured points is always higher in areas where scanning beams are closer to the axis of inclination.

Thrun et al. and Zhao et al. are using multiple static mounted 2D laser rangefinders which are oriented in different axis and the planar data are merged into 3D model based on the current pose of the robot [11][15]. Thus creating a 3D model requires movement of the robot and the errors which arise from the measurements of vehicle's odometry are also projected into the 3D model. One scanner is oriented horizontally and another vertically, Zhao et al. use two additional scanners shifted by 45° from the previous two, to assure a better model with less occlusions. The pose of the robot is determined by using a 2D SLAM on the data from the horizontal scanner.

III. DATA ACQUISITION PLATFORM

A. Laser Rangefinder

In this research, the fast and robust SICK LMS sensor as in many other robotic applications is being used. The SICK LMS scanners are operating as time of flight sensors: measuring the time it takes a laser light pulse to travel to an obstacle, where it is reflected, and return back to the sensor, where it gets registered by the sensor's receiver. The pulse laser beam is reflected by an internal rotating mirror, so that a cross-cut of the space on which the range measurements are taken in equidistant angles is perpendicular to the axis of mirror's rotation. The sensor SICK LMS 200 was mainly chosen for the following paper characteristics: it is capable of data acquisition at speeds of up to 75Hz (plane scans per second), its angular range is 180 degrees and resolution at this speed is 1 degree. In slower modes of operation, its maximum angular resolution is 0.25 degree, quadrupling the time required for the measurement. The maximum distance range of the scanner is 80 meters and the systematic error at the distance of 8 meters is ±15mm. The maximum range resolution of the scanner is 1 mm. In order to use the scanner in the fastest scanning regime (75 planar scans per second), the RS-422 serial link at 500kbps was used.

Figure 1 : 2D range finders: SICK S3000, SICK S300, Hokuyo URG

B. Inclining mechanism to obtain 3D data

A DC motor was used as an inclination actuating device. The main operation regime is precise angular velocity regulation, meaning that the inclination is changing at a constant rate. Presuming that the SICK LMS sensors have a very stable and precisely determined duration of one rotation of the mirror inside the sensor (13.30ms), if we obtain an information about one single inclination angle position of the scanner ϑ_0 at a given time t=0 when distance at rolling angle ψ_0 was measured, we can compute the remaining pitch angles ϑ of each scanned point using the following equation:

$$\vartheta = n.\omega_\vartheta.t_{rot\psi}.\frac{\psi-\psi_0}{360}+\vartheta_0 \qquad (1)$$

In equation (1), ψ is the roll angle, ω_ϑ is the inclining angular velocity; $t_{rot\psi}$ is the time it takes the mirror inside the scanner to rotate around the axis , n is the number of completed turnarounds of the mirror and $\frac{\psi-\psi_0}{360}$ is the remaining fraction of the mirror's rotation.

In order to ensure such operation of the hardware, it was necessary to ensure that the DC motor has a precise incremental encoder installed and that the motor itself is powerful enough to overcome the disturbance of gravity and other influences (friction etc.). The resulting image is not in the matrix – perpendicular form presented e.g. by Wulf et al. and Nuchter et al., since the measured points are be positioned as if the scanned lines were tilted [13][5]. Therefore, data processing is more complicated, requiring computation of the already mentioned parameter: the inclination of each scanned point.

The orientation of the scanner was also an issue. The analysis performed by Wulf et al. concluded that the orientation of the scanner is very influential on where the highest density of scanned points would be [13]. In each application, this high point density area should be directed to the area of interest, which in case of this research would be in front of the robot. This would imply the rolling inclination of the scanner. Though there were other criteria which were influencing the selection of the inclination method. A very important factor was the possible field of view of the sensor while inclining it. Another aspect was the possibility of 3DOF SLAM use during the movement of the robot which requires horizontal alignment of the scanning plane with the ground. Other aspect was the possibility of accelerating extraction of certain type of object from the 3D data – either horizontal or vertical structures (this will become clear further, in the accelerated method description). Taking all these factors, the "pitching" inclining method was selected: the scanner's scanning plane in zero position is even with the ground and the sensor is being inclined / tilted up and down, theoretically allowing whole spherical view of the surroundings. This is though only in theory since the robot itself and the construction of the mechanism always create a field of view obstacle. The continuous inclining device is presented together with the robotic platform in Figure 2.

Figure 2: The UTAR robot equipped with a 3D ranging system

C. Robotic platforms and integration of sensors

There are different robotic platforms at the Laboratory of Telepresence and Robotics which could be used to evaluate the implemented navigation method. The robot which was primarily selected for the evaluation and research is called UTAR; it is an experimental robot in the LTR which recently went through a major hardware innovation. Its major role in the future is experimental application of new autonomous navigation methods, while the robot is robust enough to carry the heaviest sensory equipment, including multiple laser rangefinders, powerful computer and heavy batteries. It can operate in both interior and exterior environment, allowing movement in light terrain (the maximum height of an obstacle is 10cm). Its maximum velocity is 2km/h, total weight is 50kg and it is powered by two 200W Portescap DC motors equipped with a harmonic transmission. The drivers are digitally controlled, allowing readings from incremental encoders connected to both motors. It is equipped with an industrial PC and the communication is performed by a WiFi MIMO module. The robot also has various DC/DC converters integrated on the chassis, allowing the integration of various actuating and sensing devices. Therefore this robot was modified for the integration of the developed 3D ranging system. The whole system is shown in Figure 2.

IV. Navigation Method

The navigation method proposed in this paper is to be applied during robot's operation in unknown environments. As noted in the previous sections, the method is in fact a simultaneous localization and mapping method in 6DOF.

The system operates in this fashion: first, the 3D data acquisition takes place. Presuming that a model of the environment is unavailable, the first scan goes only through the preprocessing phase and it is input in the model of the environment without any rotation or translation applied. The model in our system is represented by two sub-models: 2D leveled environment model and 3D environment model. The data is primarily in the form of point clouds. In the next step, the robot changes its location and when it stops moving, the 3D data acquisition takes place again. The task is to register this new 3D data set into the existing model, thus finding the approximate location of the robot and also the displacement which took place. A block diagram of the navigation system is shown in Fig. 3. The most difficult part of the system is the registration of the 3D data set. The existing algorithm are very computational time demanding.

The method presented in this paper focuses on the implementation of such navigating system which will reduce the computation time of image registration. It focuses on better use of the structures present in the 3D images to significantly reduce the number of iterations of the ICP algorithm and also the time required to match the 3D data.

A. Core of the ICP algorithm

The Iterative Closest Point (ICP) algorithm is an iterative aligning algorithm, first proposed in 1992 by P. Besl and N.D. McKay [1]. A detailed application of this algorithm in 6DOF SLAM can be found in publications by Fraunhofer AIS institute [3]. The ICP algorithm implemented in this project is very similar to the ICP presented by the Fraunhofer AIS. The Iterative Closest Point algorithm is in fact the core of the navigation method: it registers a 3D image (scan) into another image – the model. This step is based on finding the optimal rotation and translation corresponding to the displacement t and rotation R of the robot as in equation (2), for which the matching of the new scan into the global frame of reference with previous scans is most consistent. This means that the point clouds in the new scan referring to environment areas which appear also in previous scans are as close as possible to the corresponding point clouds in the map. As already stated, this rotation and translation in fact corresponds to the change in robot's pose, thus the robot is localized in the global map of the environment after each image registration.

$$\begin{aligned} &E = f(R,t,M,D) \\ &M = \{m_1,...,m_i\}, D = \{d_1,...,d_j\} \\ &i \in N, j \in N \end{aligned} \tag{2}$$

Let's assume that we have a data set D defined in equation (2), which is partly overlapping with the existing model M.

The criteria function which is to be minimized is a function $E = f(R,t,M,D)$. In the case if ICP, this function is expressed in equation 3, where $w_{i,j}$ is 1 if the point d_j in D describes the same point as m_i in M (the points are corresponding), otherwise it is set to 0.

$$E(R,t) = \sum_{i=1}^{m}\sum_{j=1}^{n} w_{i,j}\left\|m_i - (Rd_j + t)\right\|^2 \tag{3}$$

In each iterative step, the algorithm selects closest points as corresponding and calculates the transformation (R,t) according to the minimization of $E(R,t)$.

There are four known methods for minimization of this function, easy to implement are quaternion based method and singular value decomposition method (SVD). Both are applied and evaluated by Fraunhofer AIS, in later work SVD is preferred due to its robustness and easy implementation [4].

The algorithm supposes decoupling of rotation R from translation t, which is done by using centroid points given in equation (4).

$$c_m = \frac{1}{N}\sum_{i=1}^{N} m_i \qquad c_d = \frac{1}{N}\sum_{j=1}^{N} d_j$$
$$M' = \{m'_i = m_i - c_m\}, D' = \{d'_j = d_j - c_d\} \tag{4}$$
$$i \in N, j \in N$$

The final R is calculated in equation (5), where the matrices U and V are obtained by a singular value decomposition of the correlation matrix H (7); matrix H is calculated from eq. (7). Final t is derived from equation (9).

$$R = VU^T \tag{5}$$
$$H = U\Lambda V^T \tag{6}$$
$$H = \sum_{i=1}^{N} d'_i \cdot m'_i = \begin{pmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{pmatrix} \tag{7}$$
$$S_{xx} = \sum_{i=1}^{N} d'_{ix} m'_{ix}, S_{xy} = \sum_{i=1}^{N} d'_{ix} m'_{iy} \ldots \tag{8}$$
$$t = c_m - Rc_d \tag{9}$$

The most time and computation demanding task of the ICP method is the determination of the closest points. Using brute force, this is a task of searching for the closest points gives us n^2 algorithm complexity given the number of points n. There are methods to speed up the search of corresponding points, commonly used in 3D graphics programming. Majority of them is based on the structuring of point clouds using a tree structure: using octrees, box decomposition trees and kd-trees [3]. The kd-trees are a generalization of binary trees, where every nod represents a partition of a point set to the two successor nodes.

In the ICP algorithm applied in this research, it was first necessary to compute the centroids and center the data and the model. Then a kd-tree is built for the model, using a sliding midpoint splitting rule proposed by D. Mount [16]. This is followed by the building of the H matrix using closest point queries to find matching model points for all data points. This was done in two variants: the first variant does not keep track whether the model point was already used and therefore one model point is possibly used for more data closest point matches. The validity of the match was only limited by a largest distance limit of the points (typically set to 40cm). The second variant keeps track of the already used model points and restricts the reuse of them unless the distance of the current queried point from the model is lower that for the previous match.

After all data points were queried, the updated H matrix is decomposed using singular value decomposition. From the resulting matrixes the rotation is obtained and it is tested whether it is close to ones matrix. If so, the optimization does not alter the solution any more and the ICP is finished. In case the reuse of model points was restricted, the centroids are recalculated only using the matched points. The translation is determined from the rotation and the centroids, as in (9).

B. Pre-processing of the range data and data reduction, initial estimate

The preprocessor and data reduction block have two main functions: first, the range data provided by the scanning module is filtered, reducing the Gaussian noise in the data. Using the adaptive median filter, the median is computed from k points of the surrounding area $A_{k,w}$, where w is the side length of the square area A, and the k points are the ones that have the closest range values to the evaluated point. The advantages of an adaptive median filter are that it eliminates the additive noise and effaces the image without affecting the shapes and borders in the image. Secondly, data reduction is performed, especially taking into account the fact that the robot itself and the construction of the inclining mechanism appears in each data set as an obstacle. Therefore these objects are removed from the data set.

Also, the initial estimate is performed based on the odometry of the robot and the information from inclinometers. In the preliminary evaluation of the method, the inclinometers were not necessary since the testing was performed indoor and thus the inclination of the robot did not change.

C. Leveled map extraction and 3DOF pre-estimate

In mobile robotics, the robots usually stay on the ground (unless we are operating an UAV) and therefore most of their movement happens on a plane or close to a plane. Though relying on a 3DOF SLAM using the 2D planar scanner data would only be useful in strictly one-floor indoor environment. In real environments, the robot often does change its altitude and also the inclination, though most of its movement stays within certain altitude limits. This was the main motivation and an opportunity for the ICP acceleration.

In this research, the 3D range data is first used to create a leveled map with reduced number of dimensions, extracting specific features of the environment in a time-wise not costly way. In the leveled maps, vertical structures contained in 3D data are recorded, since this is a very common characteristic for both indoor (walls, steps etc.) and outdoor (trees, plants, rocks) environments.

To detect a vertical object in the data, the sensor data is processed in the following way:

1. A 3D scan is sorted so that we obtain data sets with constant yaw angles and variable pitch and range values. This is easiest if the scanning plane at zero position was originally aligned with the horizon, so that the columns in the data have uniform yaw angle.
2. These column data sets are then sorted using Combsort sorting algorithm by their vertical distance from the robot: this distance is transformed from the Cartesian coordinates as $\sqrt{x^2 + y^2}$ or from the cylindrical coordinates as $r.\cos(\vartheta)$, where r is the range value and

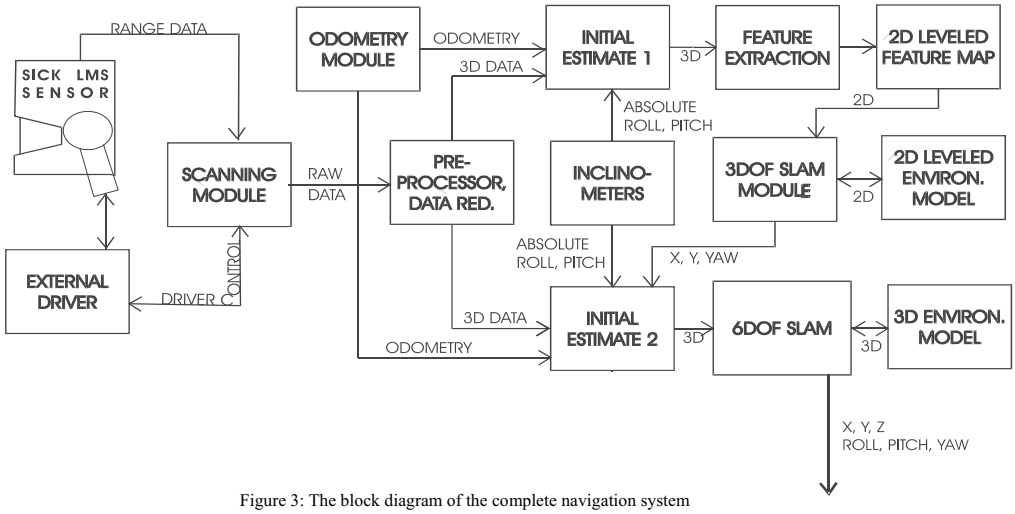


Figure 3: The block diagram of the complete navigation system

ϑ is the computed pitch angle (since the inclination is continuous through the data).

3. The subsets are scanned, looking for a required number of points in the data whose vertical distance from each other is within a specified tolerance. The number of required points is determined from the desired height of the vertical object. If such vertical object containing at least the required number of points is found, the object is added to the leveled data map.

When the leveled 2D map is constructed, it is passed to the ICP algorithm, only this time in two dimensions. A kd-tree is built for the leveled model set and then closest points to the leveled data points are queried, updating the matrix H and finally obtaining the rotation R using the SVD algorithm. This matching can only use both variants, so we can choose if model points are repetitively matched to more data points. This ICP proceeds iteratively until the rotation R is close to ones matrix. The result from this ICP registration is the rotation and translation in 3DOF (yaw angle – rotation around z axis, x and y translation).

D. Final 6DOF data registration

After the matching of the leveled map, the obtained estimate of the yaw rotation is applied to the 3D data set. Since the ICP is working with centered data sets, the application of the translation is not necessary at this point. The ICP algorithm as described earlier is applied on the 3D data, iteratively calculating the optimal rotation R, finally resulting in the full 6DOF match of the data. Then the data can be used to expand the overall model of the environment.

V. Experimental Results

For the evaluation of the proposed method, 3D range scans of an indoor laboratory environment were taken. In this

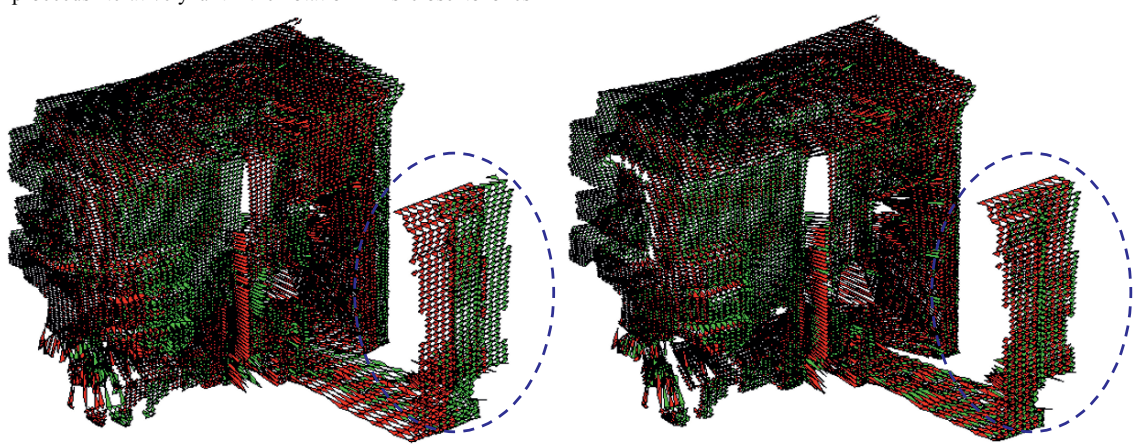


Figure 4: 3D Matched Images: standard ICP (left), ICP with unique matching of points (right)

evaluation the method is compared to the performance of a standard ICP algorithm, also taking the "unique matching" feature into account.

The following table summarizes the computing time required by the methods to match the two 3D range images together. Each of them consisted of approximately 20000 valid points and the computing machine was equipped with a Pentium M processor running at 1.86 MHz. The resulting images from the matching process are shown in Fig. 4.

TABLE I
COMPARISON OF THE EVALUATED METHODS – COMPUTING EFFORT

Selected Method	Duration [ms]	Closest pt. queries [n]
Standard 6DOF ICP SLAM	10469	356940
Leveled map accelerated standard 6DOF ICP SLAM	6406	216452
6DOF ICP SLAM with unique matching	33969	1249290
Leveled map accelerated 6DOF ICP SLAM with unique matching	8000	290544

One can observe from the table that the acceleration introduced by the leveled map introduction in the 6DOF SLAM is quite efficient in case of this indoor environment with many vertical structures. Especially when using the uniqueness of the model points' match, the computing time improved from 33969ms to 8000ms. This is especially due to the slower convergence of this version of the algorithm in early time of image registration when the two data sets are very distant. The quality of the match was not numerically compared, though by simply looking at the results in Fig. 4, one can observe that when using unique matching of data points, the method is much more precise than the standard ICP version. This is visible especially in the area highlighted by dashed circles (far from the scanner).

VI. DISCUSSION

The method for accelerating 6DOF SLAM in unknown environments was presented in this paper. It proved to be efficient in accelerating the standard though not simple ICP algorithm in indoor environments. The efficiency of the algorithm is mainly in the splitting of the difficult task into two phases, where the reduction of the number of dimensions in the first part of the SLAM process saves a lot of computing time in the next – more robust phase.

This efficiency of the algorithm though depends on its ability to detect structures in the environment – in this case vertical ones. Further evaluation and testing in different environments is subject to future research.

ACKNOWLEDGMENT

This project was supported by grant MSM0021630529 – Intelligent systems in automation and also by the Ministry of Education of the Czech Republic under Project 1M0567.

REFERENCES

[1] BESL, P. J.; MCKAY, N. D. A method for registration of 3-d shapes. In *Proceedings of IEEE Transactions on Pattern Analalysis and Machine Intelligence* 14(2), pages 239-256, Feb 1992.

[2] BIBER, P; ANDREASSON, H.; DUCKETT, T.; SCHILLING, A. 3D Modeling of Indoor Environments by a Mobile Robot with a Laser Scanner and Panoramic Camera. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'04),* Sendai, Japan, September 2004.

[3] NUCHTER, A.; LINGEMANN, K.; HERTZBERG, J.; SURMANN, H. 6D SLAM with Approximate Data Association, in *Proceedings of the 12th International Conference on Advanced Robotics (ICAR '05)*, ISBN 0-7803-9178-0, pages 242 - 249, Seattle, USA, July 2005a

[4] NUCHTER, A.; SURMANN, H.; LINGEMANN, K.; HERTZBERG, J. Semantic Scene Analysis of Scanned 3D Indoor Environments, in *Proceedings of the 8th International Fall Workshop Vision, Modeling, and Visualization 2003 (VMV '03)*, IOS Press, ISBN 3-89838-048-3, pages 215 - 222, Munich, Germany, November 2003.

[5] NUCHTER, A.; WULF, O.; LINGEMANN, K.; HERTZBERG, J.; WAGNER, B.; HARTMUT, S. 3D Mapping with Semantic Knowledge, in *Proceedings of the RoboCup International Symposium 2005*, Osaka, Japan, July 2005b.

[6] PULLI, K. Multiview Registration for Large Data Sets. In *Proceedings of the 2nd International Conference on 3D Digital Imaging and Modeling (3DIM '99)*, pages 160 - 168, Ottawa, Canada, October 1999

[7] SE, S.; LOWE, D.; LITTLE, J. Local and Global Localization for Mobile Robots using Visual Landmarks. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '01)*, Hawaii, USA, October 2001.

[8] SEQUEIRA,V.;NG, K.; WOLFART, E.; GONCALVES, J.; HOGG, D. Automated 3D reconstruction of interiors with multiple scan-views. In *Proceedings of SPIE, Electronic Imaging '99, The Society for Imaging Science and Technology /SPIE's 11th Annual Symposium*, San Jose, CA, USA, January 1999.

[9] SICK Technical Description: LMS 200 / LMS 211 / LMS220 / LMS 221 / LMS 291 Laser Measurement Systems. SICK AG – Division Auto Ident Germany. 8008970/06-2003

[10] SICK Telegram Listing: Telegrams for Operating / Configuring the LMS 2XX Laser Measurement Systems. SICK AG – Division Auto Ident Germany. 8007954/04-04-2003

[11] THRUN, S.; HAHNEL, D.; BURGARD, W. Learning Compact 3D Models of Indoor and Outdoor Environments with a Mobile Robot. In *Proceedings of the fourth European workshop on advanced mobile robots (EUROBOT '01)*, Lund, Sweden, September 2001.

[12] WULF, O.; BRENNEKE, C.; WAGNER, B.. : Colored 2D Maps for Robot Navigation with 3D Sensor Data. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, September 28-Oktober 2, 2004, Sendai, Japan.

[13] WULF, O.; WAGNER, B.. Fast 3D-Scanning Methods for Laser Measurement Systems. In *Proceedings of the International Conference on Control Systems and Computer Science (CSCS14)*, Bucharest, Romania, July 2-5, 2003.

[14] ZALUD, L.; KOPECNY, L.; NEUZIL, T.. Laser Proximity Scanner Correlation Based Method for Cooperative Localization and Map Building. In *Proc. of the 7th International Workshop on Advanced Motion Control.* pages 480-486, Maribor, Slovenia, 2002

[15] ZHAO, H.; SHIBASAKI, R. Reconstructing Textured CAD Model of Urban Environment Using Vehicle-Borne Laser Range Scanners and Line Cameras. In *Second International Workshop on Computer Vision System (ICVS '01)*, pages 284 - 295, Vancouver, Canada, July 2001.

[16] MOUNT, D. M.: Aproximate Nearest Neighbour Library. http://www.cs.umd.edu/~mount/ANN/ (22.10.2009)

A Peer-to-Peer Collaboration Framework for Multi-Sensor Data Fusion

Panho Lee, Anura P. Jayasumana, Sanghun Lim, V. Chandrasekar
{leepanho, anura, shlim, chandra}@engr.colostate.edu
Department of Electrical and Computer Engineering, Colorado State University,
Fort Collins CO 80523 USA

Abstract-**A peer-to-peer collaboration framework for multi-sensor data fusion in resource-rich radar networks is presented. In the multi-sensor data fusion, data needs to be combined in such a manner that the real-time requirement of the sensor application is met. In addition, the desired accuracy in the result of the multi-sensor fusion has to be obtained by selecting a proper set of data from multiple radar sensors. A mechanism for selecting a set of data for data fusion is provided considering application-specific needs. We also present a dynamic peer-selection algorithm, called Best Peer Selection (BPS) that chooses a set of peers based on their computation and communication capabilities to minimize the execution time required for processing data per integration algorithm. Simulation-based results show that BPS can deliver a significant improvement in execution time for multi-radar data fusion.**

I. INTRODUCTION

Recent advances in technology are introducing means to revolutionize our ability to observe, understand, and predict hazardous events by creating distributed collaborative adaptive sensing (DCAS) systems. These systems use a large numbers of *distributed* powerful sensors, such as radars, to improve spatial and temporal resolution throughout sensing area, operate the sensors *collaboratively* and *adapt* them to changing conditions in a manner that meets competing end-user needs [1]. Unlike mote-based resource poor sensor networks, which are designed to carry out one or few tasks minimizing energy expenditure, the DCAS systems are intended to serve a variety of applications and users with different requirements. Due to the massive amount of data generated by such sensors and computing intensive processing required by applications, the sensor/processing nodes are usually equipped with ample communication and computation resources.

Collaborative Adaptive Sensing of the Atmosphere (CASA) [1~5] is an example of these emerging DCAS systems. CASA is based on a network of short-range weather radars that operate collaboratively to detect and track hazardous localized weather phenomena such as tornadoes. The data generation rate at each of the radars can be several Mbps to tens of Mbps. A suite of meteorological feature detection algorithms that rely only on data from a single-radar have been used for automated identification of the hazardous weather features. To achieve improved accuracies and more specific inferences, it can be required to combine data from the multiple radar sensors using multi-sensor data fusion algorithms [2].

Client-server is one of the most popular frameworks for realizing multi-sensor data fusion. In a client-server framework, a powerful server acquires data from sensors to perform the algorithms. Although widely used, these frameworks may not be appropriate for these applications, because the demand on the server would increase in proportion to the total number of sensors; quickly overrunning server's limited capacity. To deal with the issue, peer-to-peer (P2P) framework is of interest as an alternative paradigm to the client-server architecture. In contrast to client-server frameworks, each node in peer-to-peer networks can provide bandwidth, storage and computation. For example, in file sharing P2P networks, such as BitTorrent[6], Napster[7] and Gnutella[8], each peer supplies a disk space to store the community's collection of data and bandwidth to move the data to other peers, and obtains others' resources in return. When communication cost becomes too high (e.g., in case of large multimedia content transfers), P2P networks can avoid the bandwidth limitation problem by spreading the cost over several peers. Likewise, resource-intensive sensor applications can benefit from P2P computing. In P2P frameworks, the applications can aggregate and utilize unused resources from peers over the network to achieve better performance on processing sensor data. Furthermore, the P2P mechanism also may reduce the risk of a single-point-of-failure which can be disastrous in client-server architecture.

To use a P2P framework as a feasible solution for multi-sensor data fusion, we must take the mission-critical nature of the sensor applications into account. With multi-sensor data fusion, the peers need to exchange samples to be integrated; thus the samples or data are dynamically replicated among the peers participating in the processing. Since geographically distributed sensors generate substantial volume of samples, the communication cost cannot be ignored. In addition, the large volume of generated samples must be processed with stringent time constraint, so vast amount of computation resources have to be deployed on the fly. Furthermore, in these high-end sensor networks, the peers joining the multi-sensor fusion algorithm can have a wide range of network access speeds, available bandwidth, and variability in system load. Therefore, the main problem is to coordinate a group of peers, each of which may have intensive resource demanding and high variability in performance. That is, when several sensor nodes work together on a multi-sensor data fusion algorithm they need to decide which node would be responsible for providing particular contents with acceptable performance for the multi-

This work is supported by the Engineering Research Center program of the National Science Foundation under NSF Award No. 0313747

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 207–213.

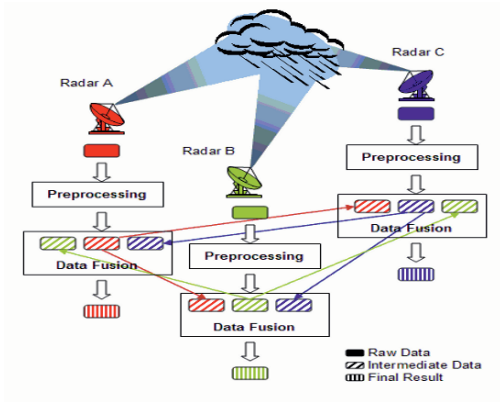


Figure 1. Multi-sensor Data Fusion

sensor data fusion algorithm. Thus, the decision must involve not only locating peers who have the desired contents, but also selecting a peer that can provide the contents within the real-time constraints. Another important issue to be considered is how to select a set of data from multiple sensors when the sampling time and intervals of the sensors are not synchronized. Because in a sensor network, each of the sensors can be assigned to a certain sensing task with different sampling schedule and mode, the synchronization of sample generation time can be a considerable task. In the presence of imperfectly synchronized sample generation, we need to provide a mechanism that notifies the peers participating in a multi-sensor date fusion algorithm when new samples are available for processing. Furthermore, a scheme is needed that selects a set of samples considering the time gap between samples and timeliness of processing.

In this paper, we present a P2P collaboration framework for CASA multi-sensor data fusion algorithms. We provide a simple data selection mechanism that addresses the problem associated with the data synchronization. We propose a dynamic peer selection algorithm, called Best Peer Selection (BPS) that allows the multi-sensor data fusion algorithm to locate desired contents in a scalable, efficient, and distributed manner, and to obtain the desired contents minimizing time. In our radar sensor network, each of the peers maintains a list of peers which are participating in the same multi-sensor data fusion process. The required contents are located by querying peers on the list. Upon receiving the query, the peers estimate the time required for providing the desired contents considering their computation and communication overhead, and respond to the query-initiator node with the estimated time. Based on the response messages, BPS can find a subset of the peers that can collectively provide the given set of contents within the time constraint. We demonstrate the feasibility and efficiency of BSP using simulations. The experimental results show that BPS can deliver a significant performance improvement, especially when the peers and the network have extremely high variability in resource availability.

The rest of the paper is organized as follows. In Section II, we introduce the system model, followed by the software architecture in Section III. Section IV presents a data synchronization mechanism. Section V details BPS, and Section VI presents the experimental results. Section VII concludes this paper.

II. TARGET SYSTEM MODEL

A high-performance radar sensor network is anticipated to consist of tens of sensor nodes connected by a combination of wired and wireless networks interconnecting the sensor nodes. The nodes run TCP/IP, and may share links with other Internet nodes. The sensor nodes are not resource constrained in terms of computation and energy compared to mote-based wireless sensor networks, and data generation rates can be several Mbps to tens of Mbps per radar sensor node. Furthermore, the sensor network is designed as a multiple end-user system. Multiple end users/applications may be present that have distinct sensing/communication/computation requirements necessary for their operations. Some of the end users/applications need to combine data from multiple sensors. Because each of the sensor nodes is allowed to conduct in-network processing and provides computation and communication resources responding to user requests, the multi-sensor data fusion can be performed collaboratively and concurrently at each of the sensor nodes [3]. Furthermore, in this real-time system, the radar sensors must be re-tasked every 30 seconds (heartbeat interval of the system) with the system goal of detecting and tracking hazardous meteorological features within 60 seconds [4].

A. Multi-sensor Data Fusion Algorithm

In our multi-sensor fusion algorithm, data acquired or received from remote sensor nodes are used as supplementary information for correcting sensing errors in the data collected by the local sensor. As illustrated in Fig 1, radar node A requires local data as well as data from node B and C to correct its sensing errors such as those due to attenuation of radar signal, while node B needs the data sampled by nodes A and C. In addition, the multi-sensor fusion algorithm usually requires a number of steps to obtain the final result. As seen in the Fig. 1, the radar data processing algorithm considered consists of two steps, preprocessing and main data fusion processing. In the first step, collected raw samples are pre-processed or sub-sampled to meet the particular requirements of the individual applications. To clarify, we define a sample as a set of radar measurements collected in a particular scanning mode (e.g., 360 degrees wide/2 elevation angle). Due to large amount of the samples collected by each of the radars, even this simple quality control, i.e., preprocessing step, of the samples takes considerable time. We had implemented the multi-sensor data fusion algorithm on a machine (3.16GHz Intel CPU, 2GB RAM) running Linux [2]. The preprocessing for the algorithm requires 1~2 seconds per sample on average. In the second step, the algorithm detects significant features by integrating the pre-processed samples. The amount of computation demand of the second step varies from 1 to 12 seconds depending on the size of the detected meteorological objects.

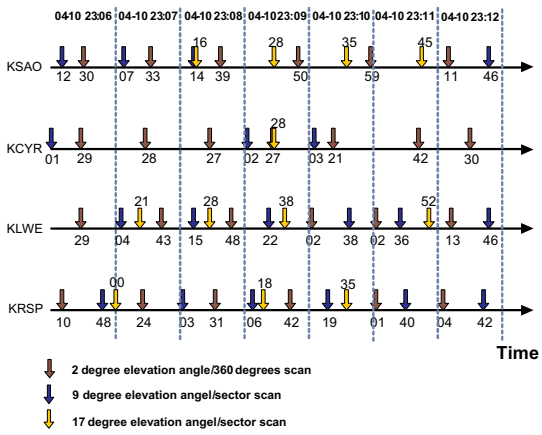


Figure 2. Sample Generation Timing

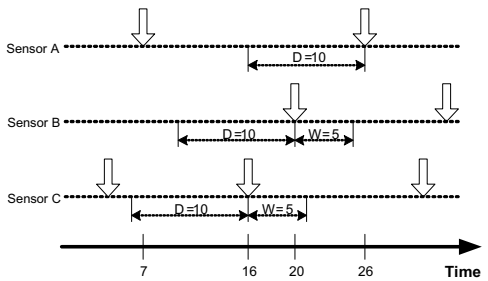


Figure 3. Data Synchronization Example.

B. *Sampling Interval*

In the sensor network, some radar scan tasks are triggered only when a particular meteorological feature is detected in the coverage area, while other tasks, such as surveillance scans at lower elevations, are triggered periodically. In addition, the various users are able to express their sensing preferences in terms of area to be scanned and sampling interval as well as the detection of meteorological features, and these preferences must be incorporated into sensing operation. The differing sensing preferences of disparate end users and reactive sensing strategies mean that it may not be possible to satisfy the needs of all end users and the required sampling interval for particular radar algorithms at the same time. Fig.2 illustrates an example of the sampling timing of CASA IP1 test-bed deployed in Oklahoma [5]. The multi-sensor data fusion algorithm considered in this paper currently requires a set of 360 degrees/2 degrees elevation angle scanned radar reflectivity samples from multiple radars.

III. NODE ARCHITECTURE

Each peer in the sensor network has an event handler, a peer manager and a resource manager. A multi-sensor data fusion application needs to continuously collect and integrate data generated from a distributed group of sensor nodes. There are a number of sensors exchanging data to be integrated, while all the required data may not be available at the time of processing. Therefore, the sensor applications have to be notified of the availability of new samples whenever new samples are generated. To provide the notification service, the event handler implements three actions:

- Delivering the data generation notifications to other nodes. A notification does not include sampled data itself, but it contains the event source, generation time, data location, and sample size.
- Providing a subscription mechanism to the radar application algorithms (RAA). Each RAA is interested in only a particular type of samples. RAAs express their interests by registering a subscription with the event handler to be notified of any forth-coming events matching the subscription.
- Informing RAAs of the generation of a sample by the local and remote sensors.

The resource manager maintains the following information at a sensor node:

- The list of application tasks the node can offer and the history of execution time for the application tasks.
- The current available bandwidth for incoming/outgoing data transfer. Every time a node communicate with another nodes, the resource manager of the node keeps track of the Exponentially Weighted Moving Average (EWMA) of incoming/outgoing bandwidth based on the time (ΔT) required to transmit/receive a sample and the size (S) of the sample as follows:

  $$IN_{EWMA} = \alpha * IN_{EWMA} + (1-\alpha) * (S_{IN} / \Delta T_{IN}),$$

  $$OUT_{EWMA} = \alpha * OUT_{EWMA} + (1-\alpha) * (S_{OUT} / \Delta T_{OUT}),$$

  To avoid too heavily being influenced by temporary bandwidth fluctuations, we currently set α to 0.8.
- The log of new sample generation events from local and remote sensors. In addition to the sample generation events, the log also contains the information about the outputs of various processing (e.g., preprocessing output, and the final results of the algorithm) and replicated contents from other peers.

As mentioned before, the multi-sensor data fusion algorithm requires obtaining samples from other sensor nodes. Obtaining data involves locating peers who have copies of the desired contents. Peers that have the desired contents are usually restricted to the peers who share same interest of processing. In order to achieve good performance for locating the contents, the peer manager maintains a list of peers who participate in the same multi-sensor data fusion processing.

IV. DATA SYNCHRONIZATION

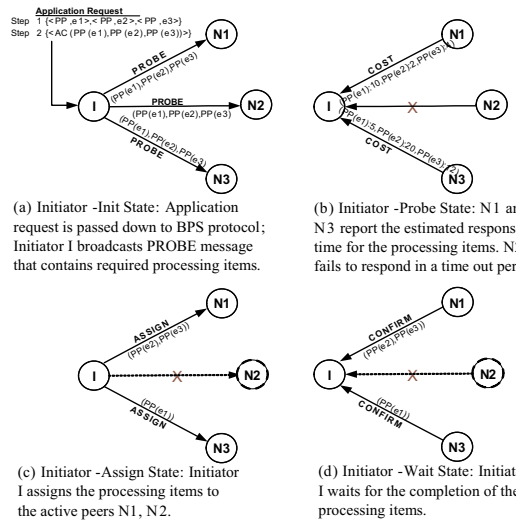
When the event handler notifies the algorithm of a new sample generation, the algorithm attempts to start the multi-sensor data fusion. First step of processing is to search the event log to find a proper set of data coming from other sensor nodes to combine them with the local data. However, as seen in Fig. 2, the sampling time of the sensor nodes is not synchronized; the algorithm may not be able to find a proper data set. Some

RAAs may want high sensing accuracy using more samples, while others may prefer timely processing to enhancing sensing accuracy; therefore RAA needs to decide whether it starts processing with the currently available samples or waits sometime to get sufficient samples for enhancing accuracy based on their application-specific processing/sensing goals. Therefore, it is required to allow the algorithms to decide their data synchronization criterion in terms of the minimum number of samples (N) from different sensors for acceptable final result, the maximum possible waiting time for samples (W) when the number of samples is not sufficient, and the maximum tolerance in sampling time difference among the samples to be integrated (D). Fig 3 illustrates an example of data synchronization for multi-sensor data integration algorithms. In the example, a particular algorithm merges samples from three sensor nodes A, B, and C with W=5, D = 10, and N = 3. At time t=16, sensor C generate a new sample, and the event handler running on sensor C posts the notification of available sample. Then the algorithm running at the node C examines the event log whether samples from other nodes are available. As seen in the figure, sensor A generated an appropriate sample at time t=7, but sensor B has not generated any samples in past 10 seconds. Thus, the algorithm has to wait until the data selection criterion (N = 3) is satisfied. At t=20, sensor B generates a new sample, and the difference 20 and 16 is less than W=5. Because all the criterion is satisfied, the algorithm can start processing at time t=20.

V. BEST PEER SELECTION PROTOCOL

After choosing a set of samples to be merged during the synchronization phase, an application starts multi-sensor data fusion by passing a request to the BPS protocol. We define the sensor nodes on which the application starts its data fusion as an *Initiator Node*. Similarly, the peers that participate in the same multi-sensor data fusion processing with the Initiator node are defined as Neighbor peers. The Initiator Node's BPS protocol has four states - ***Initiator-Init, Initiator-Probe, Initiator-Assign,*** and ***Initiator-Wait.*** Likewise, each of the neighbor peers has ***Neighbor-Init, Neighbor-Probe,*** and ***Neighbor-Proc*** states.

As illustrated in Fig. 1, because individual samples from multiple sensor nodes can be pre-processed independently, we can run preprocessing the individual samples on multiple peers. In addition, the same set of samples can be used by multiple executions of the algorithm on different nodes to correct sensing error in the data collected by individual sensors. Thus, a group of nodes interested in the same set of samples can share intermediate outputs that are obtained by preprocessing instead of raw samples. To exploit those properties of considered multi-sensor fusion processing, the application properly divides the whole process of the algorithm into several sub-preprocesses. After decomposing the whole process, the application passes down an application request that consists of the sub-processes. Each of the sub-processes is composed of a non-empty set of processing items, and each processing item is specified by a type of required processing



(a) Initiator -Init State: Application request is passed down to BPS protocol; Initiator I broadcasts PROBE message that contains required processing items.

(b) Initiator -Probe State: N1 and N3 report the estimated response time for the processing items. N2 fails to respond in a time out period

(c) Initiator -Assign State: Initiator I assigns the processing items to the active peers N1, N2.

(d) Initiator -Wait State: Initiator I waits for the completion of the processing items.

Figure 4. Best Peer Selection Protocol

and an input data set for processing. Fig. 4 illustrates how the BPS supports the multi-sensor data fusion algorithm for better performance. In the figure, the BPS protocol receives an application request which has two sub-processes: Step 1 {<PP, {e1}>, <PP, {e2}>, <PP, {e3}>} and Step 2 {<AC, {PP(e1), PP(e2), PP(e3)}>}. The first sub-process consists of 3 processing items and each of the processing items has a processing type (PP) and a sample (e.g., e1, e2, and e3) as an input for processing. The second sub-process has a single processing item and the outputs of the first sub-processes are used as the inputs for processing.

The Initiator node waits for a request from a radar application in the *Initiator-Init* state. Upon receiving the application request, the Initiator broadcasts a probing message, PROBE, to its neighbor peers. The PROBE message contains the first sub-processes (Fig. 4(a)). After broadcasting the PROBE message, the Initiator moves to the *Initiator-Probe* state, and waits for neighbor peers' response to the PROBE (Fig. 4(b)). The arrival of a PROBE message from the Initiator causes the neighbor peers to move from the *Neighbor-Init* state to the *Neighbor-Probe* state. In the *Neighbor-Probe* state, the Neighbor peers calculate the estimated time for providing the processing items specified by the PROBE message. The details of the time estimation are described in following subsection. After estimating the required time, each of the Neighbor peers responds to the Initiator Node with a COST message, which contains the estimated time for providing the desired contents. As illustrated in Fig 4(b), due to network congestion or outage or high system load, some peers may not be able to respond to the PROBE message in a certain time out period. In these cases, the Initiator node simply assumes that the Neighbor peers not responding are currently unavailable. Alternatively, the neighbor peers which can respond successfully in the time out period are regarded as active peers. When the Initiator receives the COST messages from all the neighbor peers or the

time out period is expired, the Initiator node moves from the *Initiator-Probe* state to the *Initiator-Assign* state. In the *Initiator-Assign* state the Initiator selects a peer among the active peers who has the minimum estimated time for a particular processing item based on the COST message. Once the Initiator node selects the neighbor peers for all the processing items, it assigns processing items to the selected peers by sending ASSIGN messages (Fig. 4(c)). After assigning the processing items, the Initiator moves to the *Initiator-Wait* state. In the *Initiator-Wait* state, the Initiator node waits for CONFIRM messages from the neighbor peers, which are notifications of the completion of the assigned processing items (Fig. 4(d)). When an ASSIGN message arrives at a neighbor peer, the peer moves to the *Neighbor-Proc* state, and provides the desired contents. After providing the desired contents, the neighbor peers notify the Initiator of the completion of the assigned processing items with the CONFIRM messages. After receiving the CONFIRM messages from all the neighbor peers, the Initiator peer starts the next step processing by broadcasting PROBE messages for the next step processing items. These procedures repeat until the final sub-process is done.

A. Peer Probing

In order to respond to the Initiator's PROBE message, the neighbor nodes calculate the communication and computation costs using simple estimation schemes. The details of the cost estimation algorithm are illustrated in Fig.5. The first step to the cost estimation is to determine the status of processing items. The status of a processing item can be one of the followings: ***Exist, ExistRaw*** and ***NotExist.*** If the status of a processing item is *"Exist"* at a neighbor node, then the neighbor has the output of required processing, so the Initiator node needs only the cost of communication to obtain the processing item from the neighbor node. *"ExistRaw"* means that only the raw sample exists at the neighbor; therefore, the cost of computation for required processing and the cost of communication are needed. In the case of the *"NotExist"*, the neighbor node has neither the raw sample nor the output of the processing item, so the neighbor node needs to bring the raw sample from other peers and to run required processing. Therefore, in many cases, it is not likely to assign the processing item to a neighbor node when its status of processing item is *"NotExist"*. In the BPS, the peers do not use expensive bandwidth and processing power estimation tools to determine communication and computation costs precisely. Instead, the following simplistic approach is used for the cost estimation: the resource manager of each peer keeps track of the EWMA of incoming/outgoing throughput as mentioned in Section III. Similarly, the resource manager also maintains the history of the computation time for particular types of processing. The computation time, $t_{comp}(item[i])$ for the *item[i]* can be decided by the recent computation time for the same type of processing in the history list. Likewise, we define the estimated data transfer time to acquire sample *e*, and estimated data transfer time to send *e* after the processing as $t_{in}(e)=e.size/IN_{EWMA,}$ and $t_{out}(e)=e.size/OUT_{EWMA,}$ respectively.

B. Processing Item Allocation

The cost information provided by the neighbor peers is used to assigns the processing items to the active peers. During the iterations of allocation procedure, from the unassigned processing items a new processing item is selected and assigned to an active peer that has the minimum cost for the selected processing item. After assigning the processing item to the peer, the BPS protocol adds the cost of the selected processing item to the other processing items' costs estimated by the selected peer. By updating the other items' estimated cost, the chance for assigning multiple items to the same peer can decrease. The process is repeated until all processing items are assigned to the active peers.

```
Neighbor Node Probe State
Input: Item   item[];   /* from the probe message */
COST          cost[];   /* total execution cost */
int           i, j;     /* index variable */
for each item j {
  cost[j].item = item[j];
  cost[j].total_cost = 0;
  for each data i in item[j].dataset {
    LOG log = get_ log(item[j].dataset[i]);
    if (log.status == NotExist) {
      cost[j].total_cost = t_in(item[j].dataset[i]) + t_comp(item[j]) +
t_out(item[j].dataset[i]);
    }
    else if (log.status == ExistRaw) {
      cost[j].total_cost = t_comp(item[j]) + = t_in(item[j].dataset[i]);
    }
    else if(log.status = Exist) {
      cost[j].total_cost = t_out(item[j].dataset);
    }
  }
}
```

Figure 5. Probe state algorithms for neighbor peers

VI. EXPERIMENTAL RESULT

SimGrid[10], a discrete time simulator was used in testing. The sensor nodes in our simulation sensor networks were divided into groups of maximum 4 sensor nodes for the multi-sensor data fusion. As described in Section IV, an application can decide a synchronization preference, such as the minimum number of samples (N), the maximum waiting time for samples when the number of samples is not enough(W), and the maximum tolerance in sampling time gap among the samples to be integrated (D). For the experiments, we decided D=45sec., W=5sec., and N=3. During each minute, each sensor collected 1~2 new radar samples in a particular scan mode, and each sample size was about 1MB. We set the execution time for preprocessing on a sample to 1 second. The processing times of the second step are exponentially distributed with 3 seconds mean, and we set the maximum and minimum processing time of the second step to 1 second and 12 seconds respectively. The metric we used for evaluating the performance of the BPS was response time. The response time is defined as the duration from the start of processing to the end of processing including preprocessing and second step processing.

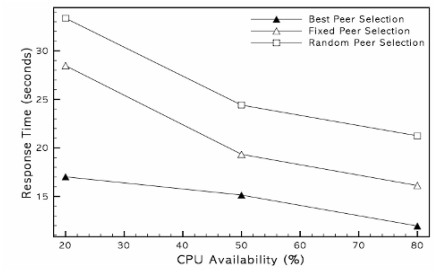
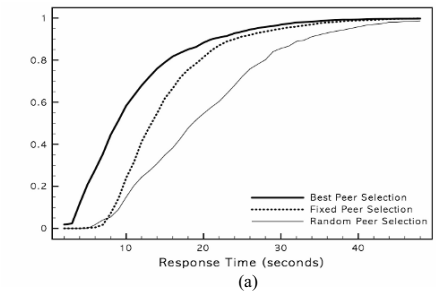
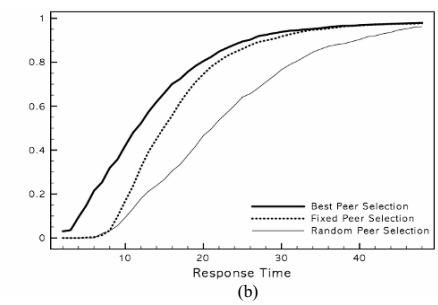


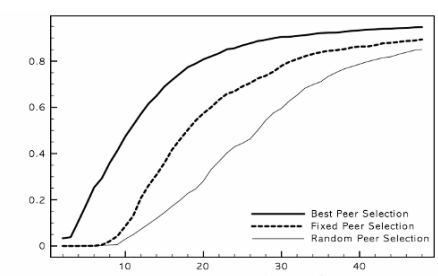
Figure 6. Average Response Time under different CPU availability.



(a)



(b)



(c)

Figure 7. CDF of response time under (a) 80%, (b) 50%, and (c) 20% CPU availability

Thus, low response time represents an improved performance. For the comparison, we also implemented two common heuristic algorithms: random and fixed peer selection algorithms. In the random algorithm, a set of peers was randomly chosen for the processing without considering the resources or contents availability at the peers. In the fixed algorithm, the peers who generated particular samples were always selected for processing the samples.

In the first set of experiments, we studied the performance of the BPS algorithm when the CPU availability was varied across the peers. In the second set of experiments, we investigated the BPS algorithm's ability to accomplish better performance in the presence of variable cross traffic in the sensor network.

A. System Load Variation

For the first set of experiments, we created a network of 12 sensor nodes. In order to simulate the CPU availability variation among the peers, we selected half of the sensors randomly, and set CPU availability to average 20%, 50% and 80% on the selected peers. The other half of the peers were assigned average 95% of CPU availability. The 50% CPU availability means that the peer can deliver only the half of its computing power to the multi-sensor data fusion algorithm. The end-to-end network bandwidth between any two sensor nodes is set to 100Mbps, thus the communication overhead for the processing is small in this set of experiments.

Fig. 6 and Fig. 7 show the simulation results for the experiments. As seen in Fig. 6, the BPS shows better results for all the conditions. The reason is that BPS selects the peers to assign the processing items considering the peers' computing time history for the same type processing but random and fixed do not. In Fig. 7, we plot the cumulative distribution function (CDF) of the response times of multi-sensor data fusion algorithm. As seen in the figures, even the selected peers is highly loaded, almost 80% of multi-sensor data fusion algorithm response times are less than 20 seconds.

B. Network Load Variation

We investigated a sensor network with 20 sensor nodes located at the leaf nodes of the network topology. We used the GT-ITM [9] to create a 100-node transit-stub graph as our underlying network topology. The bandwidths on the network links were assigned randomly as 1Gbps, 100Mbps, 10Mbps, and 4Mbps, and the network latency were set randomly to 0.5ms ~ 5ms. To simulate dynamic network conditions, we located network load generators in our sensor network. The load generators injected cross-traffic on shared links. We varied the ratio of the average cross-traffic (μ) to physical bandwidth from 0.1 to 0.9 to investigate the effect of network congestion on the response time. Plentiful evidence suggests that the network traffic being extremely variable and bursty [11,12]. We took the burstiness into account by generating the cross-traffic in a type of self-similar process known as *Fractional Gaussian Noise* (FGN) [12]. The burstiness of network traffic is usually characterized by a self-similarity parameter (Hurst Parameter, H). For the simulations, we fixed H = 0.79, and the output variance = $0.3*\mu$. Fig. 8 shows the average response time achieved by BPS, fixed, and random under different amount of cross-traffic. The results show that BPS outperformed the other two heuristics under all network conditions.

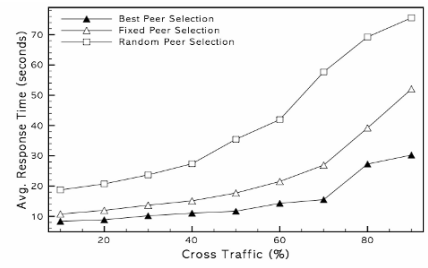
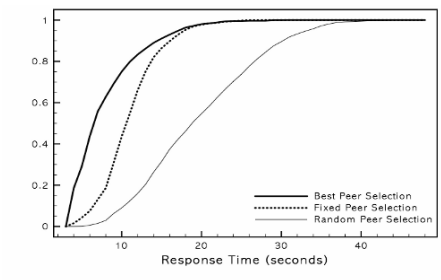
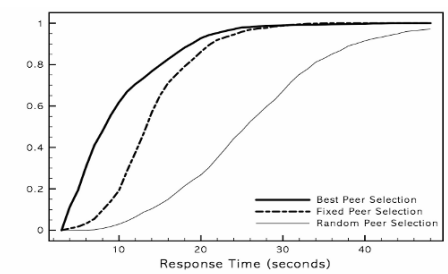


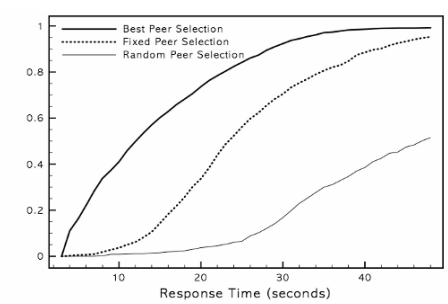
Figure 8. Average Response Time under different cross-traffic.



(a)



(b)



(c)

Figure 9. CDF of response time under (a) 10%, (b) 40%, and (c) 90% network cross-traffic

This is because in the BPS a node always chooses those peers which can provide the desired contents with the minimum cost. While collaborating on a multi-sensor data fusion algorithm, a sample from a bandwidth-poor peer can be replicated by a bandwidth-rich peer. Once the replication is done, other peers can retrieve a copy of the desired contents from the bandwidth-rich peer in the BPS instead from the bandwidth-poor source node. Fig. 9 plots the CDF of response time that BPS, fixed and random heuristics obtain. Compared to the other heuristics, the majority of the response times are smaller than that of the other heuristics.

VII. CONCLUSION

In this paper, we proposed a P2P collaboration framework for multi-sensor data fusion in resource-rich radar sensor networks. We proposed a simple data synchronization mechanism and peer selection scheme to coordinate peers for multi-sensor data fusion applications. We have implemented a simulation and our initial simulation results illustrated the effectiveness of the proposed framework and heuristic algorithm. P2P networks could provide an effective framework for the deployment of such real-time multi-sensor data fusion services. Our future plans include implementing a prototype of our framework and test it in a real test-bed environment. We will also investigate how to include other desirable feature such as fault tolerance into our framework.

REFERENCES

[1] McLaughlin, D.J., Chandrasekar, V., Droegemeier, K., Frasier, S., Kurose, J., Junyent, F., Philips, B., Cruz-Pol, S., and Colom, J. "Distributed Collaborative Adaptive Sensing (DCAS) for Improved Detection, Understanding, and Prediction of Atmospheric Hazards," in Proc. of AMS IIPS for Meteorology, Oceanography, and Hydrology, American Meteorological Society (AMS), 11.3, Jan 2005.

[2] Lim, S., Chandrasekar, V., Lee, P., Jayasumana, A. P., "Reflectivity Retrieval in a Networked Radar Environment: Demonstration from the CASA IP-1 Radar Network," Proc. of IGARSS07, Barcelona, Spain. Jul. 2007

[3] Donovan, B., McLaughlin, D., Kurose, J. V. Chandrasekar "Principles and Design Considerations for Short-Range Energy Balanced Radar Networks", Proc. of IGARSS05, Seoul, Korea, pp. 2058-2061, Jul. 2005.

[4] Zink, M., Westbrook, D., Abdallah, S., Horling, B., Lyons, E., Lakamraju, V., Manfredi, V., Kurose, J. and Hondl, K. "Meteorological Command and Control: An End-to-end Architecture for a Hazardous Weather Detection Sensor Network," ACM Mobisys Workshop on End-end Sense-and-response Systems, pp. 37-42, Jun. 2005.

[5] Brotzge, K., Brewster, B., Johnson, B., Philips, M. Preston, D. Westbrook and M. Zink, "CASA'S First Test Bed: Integrative Project #1," AMS 32nd Conf. Radar Meteor., Albuquerque, NM., 2005

[6] BitTorrent. http://www.bittorrent.com/

[7] Gnutella. http://gnutella.wego.com.

[8] Napster. http://www.napster.com

[9] Calvert L. K., Doar, M., Zegura, E., "Modeling Internet Topology", IEEE Communications Magazine, vol. 65, no. 6, Jun. 1997. pp. 160-163

[10] Legrand, A., Marchal, L., Casanova, H., "Scheduling Distributed Applications: the SimGrid Simulation Framework," Proc. of CCGrid 2003. pp. 138-145, May 2003

[11] Leland, W.E. et al., "On the Self-Similar Nature of Ethernet Traffic" IEEE/ACM Trans. Networking, vol. 2, no. 1, 1994, pp. 1–15.

[12] Paxson, V. "Fast, Approximate Synthesis of Fractional Gaussian Noise for Generating Self-similar Network Traffic," Computer Communication Review, vol. 27, pp. 5-18, Oct. 1997.

Software Implementation of Explicit DMC Algorithm with Improved Dependability

Piotr Gawkowski*, Maciej Ławryńczuk†, Piotr Marusak†, Piotr Tatjewski† and Janusz Sosnowski*
*Institute of Computer Science, Warsaw University of Technology, Warsaw, Poland
{P.Gawkowski, J.Sosnowski}@ii.pw.edu.pl
†Institute of Control and Computation Engineering, Warsaw University of Technology, Warsaw, Poland
{M.Lawrynczuk, P.Marusak, P. Tatjewski}@ia.pw.edu.pl

***Abstract*- The paper presents an approach to improve the dependability of software implementation of the explicit DMC (Dynamic Matrix Control) Model Predictive Control (MPC) algorithm. The investigated DMC algorithm is implemented for a control system of a rectification column – a process with strong cross-couplings and significant time delays. The control plant has two manipulated inputs and two outputs. The fault sensitivity of the proposed implementation is verified in experiments with a software implemented fault injector. The experimental results prove the efficiency of proposed software improvements.**

I. INTRODUCTION

Faults that appear during system operation may be critical for applications implemented in the microprocessor devices resulting in logical errors and application failure [1, 10]. It is particularly critical in control systems where faults affecting the code implementing the control algorithm can negatively influence economic efficiency or even safety of the process [3, 4]. Thus, it is extremely important to make the algorithms as robust to such faults as possible. This paper studies some basic software methods of dependability improvement of software implementation of the explicit version of Dynamic Matrix Control (DMC) Model Predictive Control (MPC) algorithm applied to a rectification column [23]. In the research the software implemented fault injector adapted to reactive applications is used [5] as during preliminary research of this control system dependability described in [3].

Model Predictive Control algorithms are widely used in the industry thanks to the advantages they have [12-14, 17, 21, 22]. They can offer better performance than the one offered by classical control algorithms, especially for processes with multiple inputs and multiple outputs (MIMO) and with significant time delay, as in the rectification column used as an example for the tests. The advantages of MPC algorithms result from the way the control action is derived. It is made using a model of the control plant. Thanks to this fact, MPC algorithms contain a decoupling mechanism which is very important for MIMO processes with strong cross-couplings. One of the MPC algorithms, often applied in practice is Dynamic Matrix Control DMC that is based on an easy to obtain step-response model [2, 17, 21].

In the next section the explicit DMC control algorithm is described. Section 3 presents software fault injector and experiment set-up. Then, section 4 contains description of software-implemented improvements toward dependability of the explicit DMC algorithm. Illustrative experiment results are presented and discussed in section 5. The paper is shortly summarized in the last section.

II. EXPLICIT DMC ALGORITHM

During the derivation of control action by MPC algorithms future behaviour of the control system is taken into consideration many time instants ahead. The prediction of process behaviour is made using a dynamic model of the control plant. The future control values are usually derived in such a way that the predicted behaviour of the control algorithm minimises the following performance index [13, 17, 21]:

$$J = (\bar{\boldsymbol{y}} - \boldsymbol{y})^T \cdot \boldsymbol{\Psi} \cdot (\bar{\boldsymbol{y}} - \boldsymbol{y}) + \Delta \boldsymbol{u}^T \cdot \Lambda \cdot \Delta \boldsymbol{u} \tag{1}$$

where $\boldsymbol{\Psi}$ is a weighting matrix of dimensionality $n_y N \times n_y N$, $\boldsymbol{\Lambda}$ is a weighting matrix of dimensionality $n_i N_u \times n_i N_u$, N and N_u denote prediction and control horizons, n_y, n_i denote the number of outputs and inputs, respectively; $\Delta \boldsymbol{u}$ is a vector of dimensionality $n_i N_u$ composed of the future increments of control values $\Delta u^j_{k+i|k}$ (for the j^{th} input, calculated at the k^{th} time instant for $(k+i)^{\text{th}}$ time instant), $\bar{\boldsymbol{y}}$ is a vector of dimensionality $n_y N$ composed of set–point values $\bar{y}^j_k$ for the j^{th} output, $\boldsymbol{y}$ is a vector of dimensionality $n_y N$ composed of output values $y^j_{k+i|k}$ predicted for the $(k+i)^{\text{th}}$ sampling instant at the k^{th} sampling instant using a control plant model. These predictions are calculated from a dynamic model of the process. The DMC algorithm is based on a step response model. Thus, if the performance function (1) is minimised without constraints then the optimal solution can be calculated using the following formula:

$$\Delta \boldsymbol{u} = \boldsymbol{K} \cdot (\bar{\boldsymbol{y}} - \boldsymbol{y}^0) \tag{2}$$

where

$$\boldsymbol{K} = (\boldsymbol{G}^T \cdot \boldsymbol{\Psi} \cdot \boldsymbol{G} + \boldsymbol{\Lambda})^{-1} \boldsymbol{G}^T \cdot \boldsymbol{\Psi} \tag{3}$$

$\boldsymbol{G}$ is a matrix of dimensionality $n_y N \times n_i N_u$ called a dynamic matrix composed of the elements of the process step response.

Because only the elements $\Delta u^j_{k|k}$ from the vector $\Delta \boldsymbol{u}$ are used (applied to the process) and then the procedure is repeated

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 214–219.

in the next sampling instant, the DMC control law (2) can be formulated:

$$\begin{bmatrix} \Delta u^1_{k|k} \\ \Delta u^2_{k|k} \\ \vdots \\ \Delta u^{n_i}_{k|k} \end{bmatrix} = \mathbf{K^e} \begin{bmatrix} \overline{y}^1_k - y^1_k \\ \overline{y}^2_k - y^2_k \\ \vdots \\ \overline{y}^{n_y}_k - y^{n_y}_k \end{bmatrix} - \sum_{j=1}^{D-1} \mathbf{K^u_j} \begin{bmatrix} \Delta u^1_{k-j} \\ \Delta u^2_{k-j} \\ \vdots \\ \Delta u^{n_i}_{k-j} \end{bmatrix} \quad (4)$$

where $\boldsymbol{K}^e$ is a matrix of dimensionality $n_i \times n_y$ and $\boldsymbol{K}^u_j$, $j=1,\ldots,D-1$ are matrices of dimensionality $n_i \times n_i$ composed of controller coefficients. The matrices $\boldsymbol{K}^e$, $\boldsymbol{K}^u_j$, and the resulting control law can be calculated off–line. The detailed description of the explicit DMC algorithm derivation can be found in [16, 21].

III. Experiment Set-Up

The concept of the Software Implemented Fault Injector (SWIFI) is based on the software emulation of a fault during the run-time of the application under test. In this research FITS fault injector is used [6, 19]. It uses standard Win32 *Debugging API* to control the execution of the software application under tests. In the following the process controlled by the analysed control algorithm (DMC) is described, instrumentation (facilitating further analysis) of the tested application is introduced, then the fault insertion policy, and finally, applied result qualification is given.

A. MIMO Process Description

The process is a rectification column with two manipulated (inputs) and two controlled (outputs) variables shown in Fig. 1. It is described by the continuous-time transfer function model [16, 23] (time constants in minutes):

$$\begin{bmatrix} Y^1(s) \\ Y^2(s) \end{bmatrix} = \begin{bmatrix} \dfrac{12{,}8}{16{,}7s+1} & \dfrac{-18{,}9e^{-4s}}{21{,}0s+1} \\ \dfrac{6{,}6e^{-8s}}{10{,}9s+1} & \dfrac{-19{,}4e^{-4s}}{14{,}4s+1} \end{bmatrix} \cdot \begin{bmatrix} U^1(s) \\ U^2(s) \end{bmatrix} + \begin{bmatrix} \dfrac{3{,}8e^{-4s}}{14{,}9s+1} \\ \dfrac{4{,}9}{13{,}2s+1} \end{bmatrix} \cdot U^3(s) \quad (5)$$

where the controlled variables are: y^1 – methanol concentration in the distillate (the top product), y^2 – methanol concentration in the effluent (the bottom product), the manipulated variables are: u^1 – flow rate of the reflux, u^2 – flow rate of the steam into a boiler, u^3 is feed flow rate (a disturbance). All process variables are scaled.

For the considered rectification process the explicit DMC algorithm is designed: the sampling period T_p=1 min is assumed, the dynamics horizon is equal to the prediction horizon D=N=100, the control horizon N_u=50, the values of weighting coefficients are: $\psi^1=\psi^2=1$, $\lambda^1=\lambda^2=10$. The simulation horizon is 300 discrete time-steps. The structure of the control system with the explicit DMC algorithm is shown in Fig. 2. The simulation scenario is as follows:

- the process is driven to a given set-point (y^1=1, y^1=0),
- at sampling instant 30 the change in the feed stream flow rate (u^3) is introduced (from 0 to 0.1),
- at the instant 140 another change in u^3 is made (from 0.1 to –0.05).

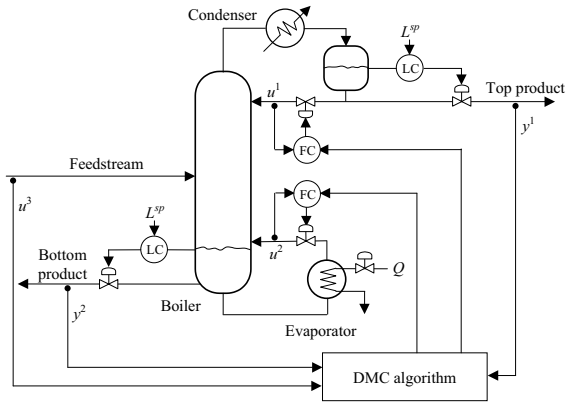


Fig. 1. Rectification column control system structure

B. Code Instrumentation

FITS disturbs directly the tested application only within so-called testing areas [5]. Testing areas limit the scope of disturbances only to the selected parts of the application. Here, the code of the controlled process model is added. The parts of the tested application disturbed during the experiments (dashed box) as well as process models (not disturbed) are marked in Fig. 2. The tested application is also instrumented to send some measures (e.g. related to internal variables values, output signal deviations) to the fault injector using user-defined messages (collected by FITS) [5].

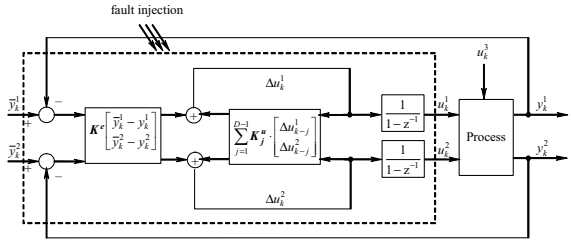


Fig. 2. Structure of the control system with the explicit DMC controller

C. Fault Injection Policy

FITS simulates faults by disturbing the running application. In this study the single bit-flip faults within CPU and FPU registers, application's data and machine instruction code are considered. Faults are injected pseudorandomly in time of program execution and space (bit position within disturbed resource, distribution over application's memory). Such a fault model well mimics Single Event Upset (SEU) effects [9, 10, 18].

D. Qualification of Experimental Results

The correctness qualification of results produced by the application under test is more complicated in case of applications related to control systems [18] than in case of

simple calculation-oriented ones. Control algorithms require complex analysis of the controlled process behaviour. The standard factor ISE (Integrated Sum of Squared Errors) is used as a measure of result (y^1, y^2) correctness. The reference ISE value (obtained during referential execution – non-faulty) is 2.53 (due to delayed response and feed stream disturbances). The whole experiment is conducted by FITS automatically. At the end of the experiment synthetic (aggregated) results for each fault location are given. In general, 4 classes of test results are distinguished:

- C: correct behaviour (ISE<5),
- INC: incorrect (unacceptable) behaviour (ISE≥5),
- S: test terminated by the system due to un-handled exception,
- T: timed-out test.

Analysis of fault effects requires detailed information upon the faults injected and the application behaviour. FITS provides details about every test (simulated fault injection). Hence, manual replay of the whole test execution can be done. Moreover, all the events and user messages occurring during the test are recorded. The tested application is instrumented to save its outputs (here simulation results, i.e. a set of control signals in subsequent sampling instants) into separate files for each test (file names are managed by FITS). This gives a possibility for post-experiment analysis of fault effects in the correlation with the injected fault and observed behaviour for each test.

IV. Considered Software Implementations

A. Classical Implementation

Classical software implementation of the DMC algorithm for considered rectification column is made in the C language. Within its body several calculation steps can be distinguished:

- Calculation of increments of manipulated control variables values (according to the equation 4, e.g. control errors calculation) – the major part of the algorithm,
- Saving the calculated values of increments into the historical vector of increments (needed for further processing),
- Calculation of manipulated control variables values for the next iteration (integration of previous values with the increments calculated in current iteration).

Implementation of the control algorithm takes 124 machine instructions (405 bytes of the static code). The algorithm needs execution of 1020000 instructions for the whole simulation horizon (300 discrete sampling instants). The application (and its other, further described versions) is compiled with the MS Visual C++ 2005 compiler. The classical implementation is compiled with the default compiler's options (in *Release* configuration). In particular, the /EHsc parameter is provided (enabling C++ exception handling without support for native Win32 Structured Exception Handling - SEH) even though there is no exception handling statements in the source code. As the static and dynamic profile is different, the distribution of mnemonics in the static code as well as in the executed stream (dynamic) is presented in Fig. 3.

It is worth noting that floating point instructions constitute 38% of the code in the static code while dynamically they are executed in 54.8% of time. Moreover, instructions organising computational loops in the DMC implementation (*sub*, *test*, *add*, *jnz*, *jg*) take another 38%. Hence, the application is strongly computational with high degree of FPU utilisation. Nevertheless, the instruction set used is rather limited (shown on Fig 3). On the other hand, the activity ratio for CPU resources is high (98, 94, 80, 98, 97, and 81 % for EAX, EBX, ECX, EDX, ESI and EDI, respectively) [18].

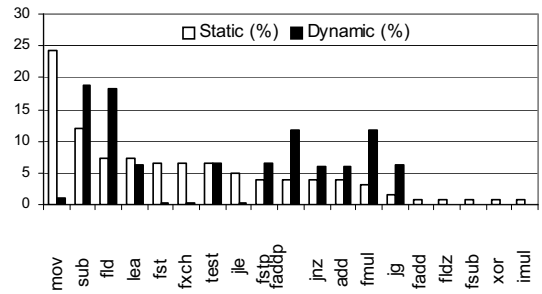


Fig. 3. Distribution of DMC mnemonics in classical implementation (sorted by static %)

The dependability of such classical software implementation is examined and discussed in [3]. Here, some software improvements are introduced. One can expect, that in the DMC case some simplifications in the fault hardening implementations can be applied without noticeable aggravation of dependability and performance (i.e. rare errors on the controlled process inputs are not critical for its behaviour). Hence, two hardened software versions are considered and the impact of different sets of parameters devoted to exception handling of the used compiler is presented.

B. Hardned Implementations

The idea of the first hardened version considered (further denoted as H1) is based on triplication of the DMC calculation (executed sequentially). Only the first step of the classical implementation is triplicated. Then, the voting over obtained control variables increments propositions takes place (before the second and third steps according to the classical implementation). The voting is based on the majority with the preconfigured threshold of acceptable differences between two propositions. If no majority is found, the arithmetic average value is calculated. After voting the second and the third step of the classical algorithm takes place. In H1 there are no exception handling statements in the code. Nevertheless, compilation parameters are the same as in the classical version. Such implementation consists of 463 machine instructions (1278 bytes of the static code) and requires the execution of 2289794 instructions to complete the task.

The second hardened implementation uses standard C++ exception handling statements (*try/catch(...)*) to capture exceptions occurring during the execution of the DMC body (the same one as in the classical implementation). In the exception-handling block the FPU unit is reset (with *_fpreset* function) and the control variable value is left unchanged from the previous control iteration. In other words, the handling is very simple. Such implementation is firstly examined with the default compiler's options: /EHsc and without support for FPU exceptions (further referred as H2a version). Then, another version (H2b) is produced by compiling the same hardened code with /EHa (support for SEH enabled) and /fp:except flags (support for FPU exceptions enabled). It is interesting that setting those parameters increases the number of executed instructions within the DMC from 1020000 to 1940400 while the static code size grows only from 405 bytes (124 instructions in H2a) to 461 bytes (125 instructions in H2b). The impact of compilation options needs further research as the dependability of executables differs a lot (discussed in the next section). Fig. 4 presents the profile of H2b mnemonics distribution as the profile of the H2a is similar to the classical implementation. It is interesting, that in the case of H2b the differences between the static and dynamic profile are much lower then in classical implementation of the DMC.

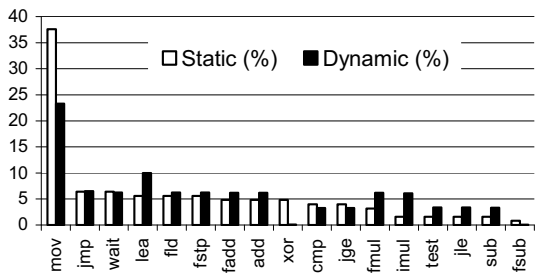


Fig. 4. Distribution of DMC mnemonics in H2b version (sorted by static %)

V. EXPERIMENTS

Faults in CPU and FPU registers, data area of the application, executed instruction stream, and static code image are considered. For each fault location approximately 1000 disturbed executions are investigated (single fault injected in each application execution).

The dependability of the classical implementation and impacts of faults within the DMC on the rectification column is presented in details in [3]. This paper focuses on the comparison of dependability improvements with simple software fault tolerance solutions. The application behaviour discussion is very limited. Fig. 5, 6, and 7 present three examples of results of single bit faults disturbing faddp instruction at different time instants. Plots of the application outputs (y1, y2) over the DMC iteration number are given. For reference the undisturbed simulation results are also shown, i.e. the golden run (solid lines in Fig. 5,6, and 7).

It is worth noting that overwhelming majority of incorrect behaviour relates to very high ISE values (higher than 1000). It means that the control values errors cumulates. On the other hand it gives the possibility to easily detect such big deviations using additional diagnostic subroutines.

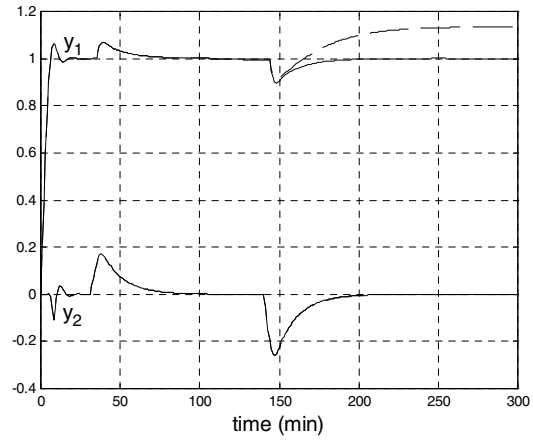


Fig. 5. Single bit inversion within the *faddp* instruction disturbs the top product composition. Fault injected at the 101st sampling instant; ISE=4.25; golden run responses – solid line, fault injected responses – dashed line.

The result of those faults varies on the control signal and process states. For instance, the considered fault injected at the 101st sampling instant in ISE=4.25 (Fig. 5), at the 170th – ISE=2.58 (Fig. 6). Hence, it disturbs the top product composition. More critical situation is illustrated in Fig. 7. Single bit inversion (at 44th sampling instant) within the same instruction as described above destabilises the process (ISE=4.40).

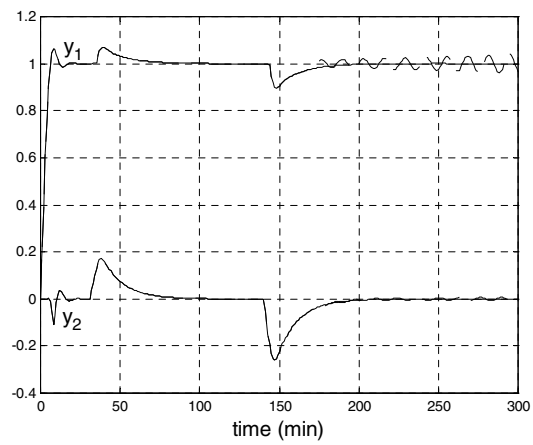


Fig. 6. Single bit inversion changes the destination operand of *faddp* instruction from st(2) to st(3). Fault injected at the 170 sampling instant; ISE=2.5831; golden run responses – solid line, fault injected responses – dashed line.

The consequences of an injected fault depend on the state of the control system. In the case the fault is injected when the system is in a steady state the responses are practically the same as those obtained in the golden run; the control errors for both outputs, at the end of simulation, are close to zero. In the cases when the fault is injected earlier (Fig. 5 and 7) the control system could not properly compensate the change in disturbance u_3. In both cases the steady state control error on the first output y_1 is nonzero. However, in the discussed case the second output y_2 is robust to the fault that occurred. Nevertheless, in the cases presented in Fig. 6, 7 the negative influence of the faults are visible on both outputs (oscillations occurred). In the second case (Fig. 7) it is clear that the system is unstable, its behaviour is unacceptable.

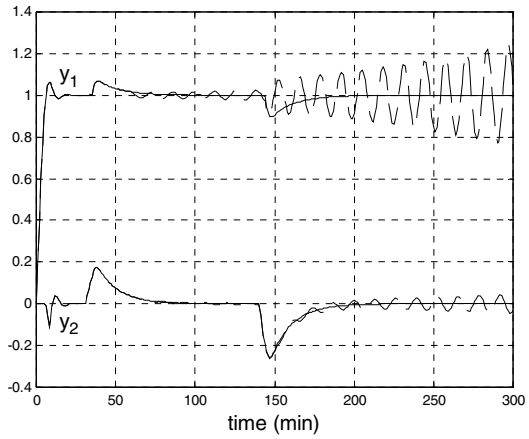


Fig. 7. Single bit inversion within the *faddp* instruction destabilises the process. Fault injected at the 44th sampling instant, ISE=4.40; golden run responses – solid line, fault injected responses – dashed line.

The summary of results (according to categories described in Section 3) is presented in Fig. 5. In each fault location, the group of bars corresponds with the results of the considered application version. Starting from the left side in each group it is classical, H1, H2a and H2b implementation respectively.

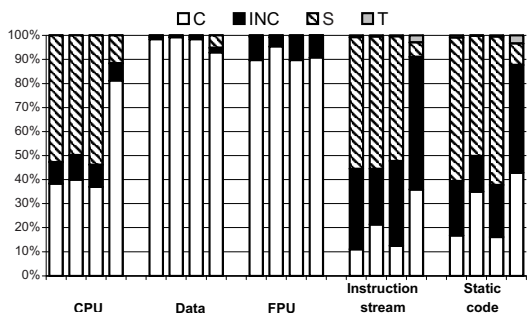


Fig. 8. Summary of experimental results

Data area comprises mostly algorithm parameters and only a few data memory locations store critical variables. Moreover, due to a large number of parameters (400) disturbing a single parameter usually has low impact on the results. Hence, faults in data area are tolerated in a large extent (C>90%). The high degree of FPU robustness could also be astonishing. However, past experience shows that the FPU is rarely used hard [10] (e.g. only few FPU stack locations used simultaneously). This results in overall low fault sensitivity of the FPU. Nevertheless, there are some very sensitive locations within the FPU (e.g. control registers).

The most fault sensitive resource of the DMC controller is its code. Hopefully, software triplication in H1 applied at the source code level to provide fault robustness almost doubled the percentage of correct executions (category C of experiment result) in case of faults disturbing code locations (executed instruction stream and the static code image as well). But without the exception handling mechanism embedded, the redundant code does not have the chance to accomplish computations, as in over 50% of test cases the injected fault causes the unhandled exception (terminating immediately the application).

The most interesting is the behaviour of the H2a version. It is surprising that by default the programmer would not obtain the expected improvement (exception handling) as the default compiler settings are not necessarily suitable for the solution. Moreover, enabling the FPU exceptions (which is not the default in VC++) seems not to solve the problem as those and many other exceptions (e.g. mostly Access Violation cases) can be only handled in the code using the Win32 native Structure Exception Handling mechanism (SEH). Presented results prove that the only /EHa compiler's parameter assures proper handling of these exceptions. The improvement is extraordinary in case of faults located in the CPU registers (over 80% of correct executions). The higher percentage of correct executions is observed also in the case of faults located in the code (twice higher than in H1 version). However, at the same time the percentage of incorrect executions (INC category) rises at the expense of unhandled exceptions in other implementations. The increase of percentage in C and INC categories is typical for simple fault tolerant implementations. Here, the calculation continues without any, even basic, fault detection or fault recovery. This drawback can be minimised, for instance, by including error detectors inside of the code, redundant code and the most important data instances or other fault recovery – not present in the considered implementations. It is clear that redundant code without the integration with SEH mechanisms is also not sufficient.

The strategy of exception handling by simply applying the previous control values for next sampling instant is very efficient in case of locations that keep temporary values, e.g. CPU registers. The considered faults in the code locations (executed instruction stream as well as the static code image) are designed to mimic mainly memory location faults. Faulty code remains unchanged as long as it is reloaded from

persistent storage (e.g. disk, EPROM, flash RAM). The solution H2b would be very efficient in case of faults located in the instruction cache as the faulty code could be recovered while flushing the cache or reading the code from fault robust location.

VI. Conclusions

The paper studies dependability of software implementation of the explicit DMC algorithm applied to a rectification process. The process has two manipulated variables, two controlled variables and significant time-delays. Dependability is examined using software implemented fault injector.

Results of the experiments carried out clearly indicate that the well-known and widely used formulation of the DMC algorithm is susceptible to hardware faults. It is particularly important in case of industrial processes such as the considered rectification column because faults are likely to lead to undesirable behaviour of the process. More specifically, in the least difficult situation a fault can disturb composition of the products whereas in the most dangerous case it can destabilise the system. Technological and financial consequences of faults are of fundamental importance. Energy losses and unacceptable compositions of the products (which means that the product cannot be sold) are just two examples of such situations.

Presented research proves that the software implemented fault handling techniques can significantly improve the dependability of the DMC. However, to achieve high degree of fault robustness the integration of fault detection, recovery and system detection mechanisms such as exceptions is needed. Moreover, the evaluation of the final binary code as well as the deep development tool features investigation is crucial. Observed big differences in fault susceptibilities of considered DMC releases are easily discovered during the fault injection experiments.

In order to increase fault robustness of software implementation of the DMC algorithm, a few techniques can be applied. In particular, current research embrace DMC algorithm based fault detection. The most critical situations (technologically unacceptable control values leading to instability) may be detected by an additional diagnostic subroutine and handled by redundant code.

Acknowledgment

P. Gawkowski and J. Sosnowski realised this research within grant 4297/B/T02/2007/33 from Polish Ministry of Science and Higher Education.

References

[1] A. Benso, P. Prinetto, *Fault Injection Techniques and Tools for Embedded Systems Reliability Evaluation*. Kluwer Academic Publishers, 2003.

[2] R. Cutler, B. Ramaker, "Dynamic matrix control – a computer control algorithm," *AIChE National Meeting*, Houston, 1979.

[3] P. Gawkowski, M. Ławryńczuk, P. Marusak, J. Sosnowski, P. Tatjewski, "Dependability of the explicit DMC algorithm for a rectification process," Proceedings of the International Multiconference on Computer Science and Information Technology, Wisła, Poland, pp. 779-788, 2007.

[4] P. Gawkowski, M. Ławryńczuk, P. Marusak, J. Sosnowski, P. Tatjewski, "Dependability of explicit DMC and GPC algorithms," Proceedings of the International Multiconference on Computer Science and Information Technology, Wisła, Poland, pp. 903-912, 2007.

[5] P. Gawkowski, J. Sosnowski, "Experiences with software implemented fault injection," Proceeding of the International Conference on Architecture of Computing Systems, VDE Verlag GMBH, pp. 73-80, 2007.

[6] P. Gawkowski, J. Sosnowski, "Analysing system susceptibility to faults with simulation tools," *Annales UMCS Informatica AI*, vol. 4, pp. 123-134, 2006.

[7] P. Gawkowski, J. Sosnowski, "Software implemented fault detection and fault tolerance mechanisms – part I: Concepts and algorithms," *Kwartalnik Elektroniki i Telekomunikacji*, vol. 51, pp. 291-303, 2005.

[8] P. Gawkowski, J. Sosnowski, "Software implemented fault detection and fault tolerance mechanisms – part II: Experimental evaluation of error coverage," *Kwartalnik Elektroniki i Telekomunikacji*, vol. 51, pp.495-508, 2005.

[9] P. Gawkowski, J. Sosnowski, B. Radko, "Analyzing the effectiveness of fault hardening procedures," Proceedings of the 11th IEEE Int. On-Line Testing Symposium, pp. 14-19, 2005.

[10] P. Gawkowski, J. Sosnowski, "Dependability evaluation with fault injection experiments," *IEICE Transactions on Information & System*, vol. E86-D, pp. 2642-2649, 2003.

[11] P. Gawkowski, J. Sosnowski, "Experimental Validation Of Fault Detection And Fault Tolerance Mechanisms," Proceedings of the 7th IEEE International Workshop on High Level Design Validation And Test, Cannes, France, 2002.

[12] M. Ławryńczuk, P. Marusak, P. Tatjewski, "Multilayer and integrated structures for predictive control and economic optimisation," Proceeding of the IFAC/IFORS/IMACS/IFIP Symposium on Large Scale Systems: Theory and Applications, Gdańsk, Poland, CD-ROM, paper 60, 2007.

[13] J. M. Maciejowski, *Predictive Control with Constraints*. Prentice Hall. Harlow, 2002.

[14] M. Morari, J. H. Lee, "Model predictive control: past, present and future," *Computers and Chemical Engineering*, vol. 23, pp. 667-682, 1999.

[15] S. J. Qin, T. A. Badgwell, "A survey of industrial model predictive control technology," *Control Engineering Practice*, vol. 11, pp. 733-764, 2003.

[16] J. Pułaczewski, "Multidimensional DMC algorithm," *Report of Institute of Control and Computation Engineering, Warsaw University of Technology*, no. 98-11, (in Polish) Warsaw, 1998.

[17] J. A. Rossiter, *Model-Based Predictive Control*. CRC Press, Boca Raton. 2003.

[18] J. Sosnowski, P. Gawkowski, A. Lesiak, "Fault injection stress strategies in dependability analysis," *Control and Cybernetics*, vol. 33, pp. 679-699, 2005.

[19] J. Sosnowski, A. Lesiak, P. Gawkowski, P. Włodawiec, "Software implemented fault inserters," Proceedings of the IFAC Workshop on Programmable Devices and Systems, Ostrava, pp. 293-298, 2003.

[20] J. Sosnowski, P. Gawkowski, A. Lesiak, "Fault injection stress strategies," Proceedings of the 4th IEEE LATW 2003 Workshop, pp. 258-263, 2003.

[21] P. Tatjewski, *Advanced Control of Industrial Processes, Structures and Algorithms*. Springer. London, 2007.

[22] P. Tatjewski, M. Ławryńczuk, P. Marusak, "Linking nonlinear steady-state and target set-point optimisation for model predictive control," Proceedings of the IEE International Control Conference ICC 2006, Glasgow, 2006, CD-ROM.

[23] R. K. Wood, M. W. Berry, "Terminal Composition Control of a Binary Distillation Column," *Chemical Engineering Science*, vol. 28, pp. 1707-1717, 1973.

Simulation Strategy of the Analog Front End for UHF Passive RFID Transponders

Qiuyun Fu, Nicolas Gay, Christian Bansleben, Wolf-Joachim Fischer
Fraunhofer Institute for Photonic Microsystems (IPMS)
Maria-Reiche-Str. 2, 01109 Dresden, Germany

***Abstract*—The analog front end is one of the key elements in passive UHF RFID transponders, accounting for such diverse operations as RF-DC conversion, voltage regulation, modulation und demodulation of the incident RF wave, and voltage limiting, among others. During the design process it is therefore necessary to accurately simulate the power up process and such functions as demodulation and backscatter modulation. However, since the input impedance of the chip varies with different operating conditions, the test bench set up poses a complex task. In addition, the antenna connected to the front end plays a key role and must be adequately modeled. This paper focuses on the simulation strategy of the analog front end for an RFID transponder fabricated in standard CMOS technology working in the UHF/SUHF range. We show how the input impedance and input power of the analog front end in its active and idle states can be precisely determined through simulation.**

I. Introduction

RFID systems have found broad application in the supply chain management, access control and public transportation [1] [2], among other fields. Basically, a RFID system consists of a reader and one or more transponders (tags). The transponder itself can be completely passive [3][4][5], semi-passive [6], or active.

The passive RFID transponder is composed of an antenna, an analog RF-front-end and a digital core. Sometimes sensors are also included, as shown in Fig. 1. The analog RF front end is located between the antenna and the digital part of the transponder. A passive transponder has no external battery. It generates its power supply directly from the incident RF signal. The antenna collects the transmitted power from the reader, inducing an RF voltage signal at the feed point of the chip. Thus, the main tasks of the analog front end can be summarized as follows:

1) Rectification of the RF electromagnetic wave into a DC signal. The latter must be then regulated in order to obtain a stable voltage supply (V_{reg}) for the digital core and other circuits.

2) Detection and demodulation of the ASK modulated incident RF signal, providing a digital bit stream (V_{dem}) for the digital backend.

3) Backscatter modulation by switching the input impedance between two reflection states as dictated by the signal V_{trdata}.

4) Generation of the power on reset signal (V_{por}) to set the digital core in a known initial state.

Such an analog front end can be fabricated employing a standard RF-CMOS technology. The front end described in this paper is based on Austriamicrosystems' 0.35µm RF-CMOS technology, featuring RF MOSFETs and passive components.

In general, once every module (rectifier, modulator, demodulator, etc.) has been designed and tested at transistor level following specifications, it is necessary to verify the chip as a whole to validate its functionality. In order to simulate the front end correctly, the power transmission between the antenna and the front end must be accounted for. According to wireless transmission theory, the maximal available power delivered by the antenna to the load (IC) takes place when the respective impedances are matched to each other. This is particularly important for passive RFID transponders where a large read out range is of primary concern. The antenna is

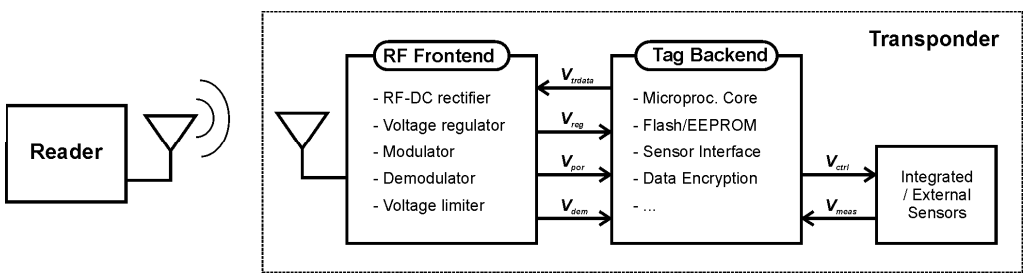


Fig.1. RFID system consisting of reader and transponder

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 220–225.

designed to match the IC's input impedance at the specified resonant frequency. However, the input impedance of the RF front end changes according to the actual input voltage level and the delivered output power. In the following text, a simulation strategy to deal with this complex dynamic situation will be described. Based on an iterative process, the impedances of the antenna and the chip are estimated. In first place, the power matching point is determined. The input impedance of the front end at the matching point is then approximately calculated through simulation, whereby the input voltage of the chip is fixed. The antenna parameters based on these initial values are estimated and included in a second simulation. After that, the input impedance is calculated once more, resulting in more accurate parameters for the antenna. The parameters thus obtained are employed in the final simulations.

II. Determination of the Power Matching Point

A simple behavioral model of the transponder is shown in Fig. 2. The antenna is represented by its Thevenin equivalent circuit. Recently, there have been discussions about limitations

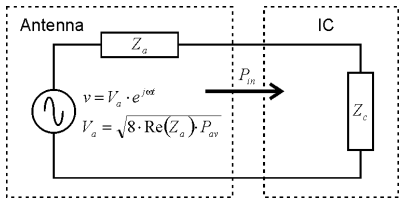


Fig.2. Antenna behavioral model

of such a model for loaded antennas in receiving and transmitting mode [7]. Nonetheless, this equivalent circuit has been successfully used to cope with various antenna problems in the past [8].

When the far field approximation is valid (i.e. the distance separating the transponder and the reader is larger than $\lambda_0/(2\pi)$), the power collected by the transponder antenna is given by

$$P_{av} = \frac{P_{EIRP}}{4\pi r^2} A_e \tag{1}$$

with

$$A_e = \frac{\lambda_0^2}{4\pi} G \tag{2}$$

where $\boldsymbol{P_{av}}$ is the maximum available power delivered by the antenna, $\boldsymbol{P_{EIRP}}$ is the equivalent isotropic radiated power (EIRP), $\boldsymbol{r}$ the readout distance, $\boldsymbol{A_e}$ the effective area of the antenna, $\boldsymbol{\lambda_0}$ the wavelength, and $\boldsymbol{G}$ the gain of the transponder antenna. From Fig. 2, when $\boldsymbol{Z_c}$ equals $\boldsymbol{Z_a^*}$ -- the conjugate complex of $\boldsymbol{Z_a}$ --, the power delivered to the IC ($\boldsymbol{P_{in}}$) is equal to $\boldsymbol{P_{av}}$.

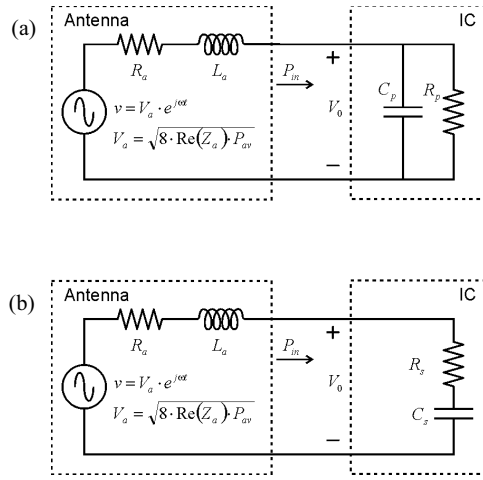


Fig.3. IC behavioral models

In general, the complex impedance of the antenna can be represented through a resistor in series with an inductor. The IC, on the other hand, shows a capacitive behavior and can thus be replaced either by a capacitor $\boldsymbol{C_p}$ in parallel with a resistor $\boldsymbol{R_p}$ (Fig. 3(a)) or a capacitor $\boldsymbol{C_s}$ in series with a resistor $\boldsymbol{R_s}$ (Fig. 3(b)). $\boldsymbol{V_0}$ is the amplitude of the voltage at the chip's feed point.

During steady operation, the capacitance $\boldsymbol{C_p}$ remains almost unchanged and can thus be considered in the model as constant. $\boldsymbol{R_p}$ can be estimated by [4],

$$R_p = \frac{V_0^2}{2P_{in}} \tag{3}$$

Considering the IC models of Fig. 3, the chip input impedance can be expressed in the following two forms:

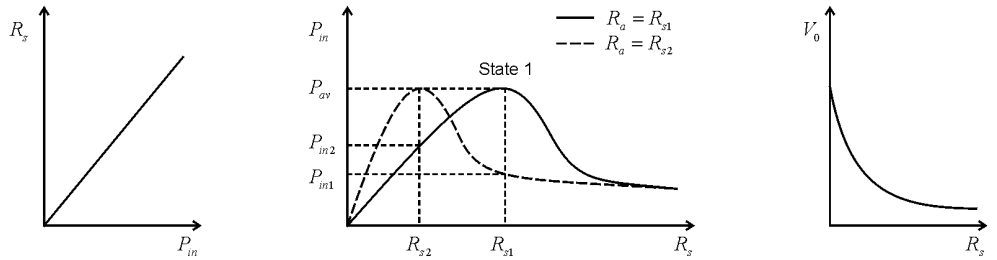
$$Z_c = \frac{1}{\frac{1}{R_p} + j\omega C_p} = R_s - j\frac{1}{\omega C_s} \tag{4}$$

where

$$\frac{1}{\omega C_s} >> R_s \quad \text{and} \quad \omega C_p R_p >> 1 . \tag{5}$$

Thus, the components in both models are related as follows

$$R_s \cong \frac{1}{\omega^2 C_p^2 R_p} \text{ and } \quad C_s \cong C_p . \tag{6}$$



(a) P_{in}-R_s characteristic (V_0 constant, eq.(3),(4))

(b) P_{in}-R_s characteristic (P_{av} constant, eq.(8))

(c) V_0-R_s characteristic (P_{av} constant, eq.(7))

Fig.4. Relations among P_{in}, R_s and V_0

From Fig. 3 and (6), it is obvious that increasing the IC power consumption results in a reduction of the resistor R_p or equivalently an increase of the series resistor R_s.

Furthermore, V_0 can be derived from Fig. 3(b) using the matching condition $\omega L_a = (\omega C_s)^{-1}$ and Eq. (5):

$$V_0 = \left| \frac{R_s - j\frac{1}{\omega C_s}}{R_a + j\omega L_a + R_s - j\frac{1}{\omega C_s}} \right| V_a \cong \frac{\sqrt{R_a}}{R_a + R_s} \frac{\sqrt{8P_{av}}}{\omega C_s} \qquad (7)$$

and from equations (3), (6) and (7), we can rewrite the IC's input power as follows:

$$P_{in} = \frac{R_s R_a}{(R_a + R_s)^2} \cdot 4P_{av} . \qquad (8)$$

After power up, the chip will progressively go through different states. Note that the power consumption (and thus P_{in}) depends on the requirements of the digital core. Figure 4 (a) shows clearly that the real part (R_s) of the IC input impedance will change with P_{in}. It was already stated that the antenna should match the IC in order to maximize the power delivered to the load. Therefore, R_s should tend to R_a. However, the impedance of the antenna does not change with P_{in}, and thus $R_a \neq R_s$. So, how do we choose the actual power matching point?

Observing Fig. 4(b), let's assume in first place that R_s is equal to R_{s1} when the IC power consumption reaches its maximal value (State 1), which is the minimal required power for correct operation of the IC. In other words, the system must reach $R_s = R_{s1}$ with $P_{in} = P_{in1}$ for correct operation. The value of R_s for any other state (State 2) is lower than R_{s1} and can be represented as R_{s2}. Two cases can be differentiated, depending on whether the final operating point is before or after the maximum:

1. Matching point at $R_{s2} = R_a$ (dashed curve): If the matching point is at State 2, P_{in} will reach P_{av} at R_{s2} and $P_{in1} < P_{av}$ at R_{s1}. It is clear, that a part of the maximal available power will be wasted, since $R_s = R_{s1}$ is found after the maximum. Moreover, because V_0 drops quickly with R_s, it will be difficult to set the transponder in a high reflectivity state (i.e. during backscatter modulation), since the voltage at its feed point is very low. In addition, the RFDC converter suffers to be shut down and the chip can not reach State 1 under a very low V_0.
2. Matching point at $R_{s1} = R_a$ (continuous curve): P_{in} will reach P_{av} at R_{s1}. Before reaching its maximum, for instance at $R_s = R_{s2}$, R_s is small and V_0 large, as indicated by Fig.4(c). A large V_0 induces an increase of P_{in} and therefore prevents R_s from decreasing. Finally, a balance will be reached and the chip will operate correctly at R_{s2}.

III. Estimation of the Input Impedance

The power matching point is selected at the peak input power (State 1). Simulations are performed to determine the chip input impedance in this state and also the value of the minimal input power required.

A. Foundation of the Simulation

The simulation conditions to determine the input impedance are as follows (see Fig.5):

1) The power consumption for the transponder's backend is represented by R_{load} and C_{load}. Their values correspond to the State 1 in Section 2.

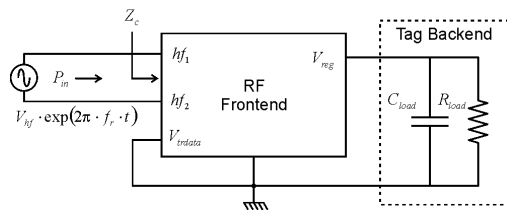


Fig.5. Test bench to estimate the IC's input impedance

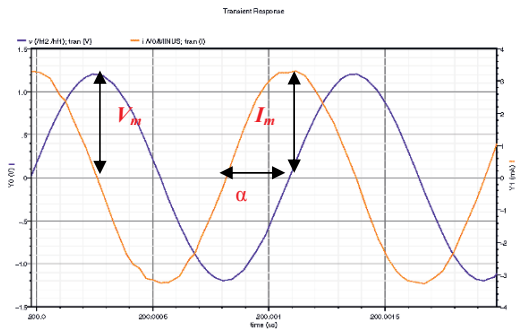


Fig.6. Simulation results for determination of chip impedance

2) The amplitude of the incident RF signal V_{hf} is selected in such a way as to let the regulated output voltage V_{reg} reach the specified output voltage (1V in our case). The RF signal has a frequency f_r (900 MHz).
3) No modulation is applied to the RF signal.
4) The backscatter modulation port V_{trdata} is grounded, i.e. the entire input power is employed in the rectification process.

B. Results Analysis

Figure 6 shows the input RF voltage V_m and the sourced current I_m., which are both used to calculate the input impedance Z_c with equation (9).

$$Z_c = V_m / (I_m \angle \alpha) \tag{9}$$

where V_m is the amplitude of the input voltage V_{hf}, I_m is the amplitude of the input current through pins hf_1 and hf_2, and α is the phase difference between both magnitudes.

From Fig. 6 we get Z_c = 10.7 – j363.5. Therefore we need an antenna with impedance $Z_a = Z_c^*$ = 10.7 + j363.5 in order to match the IC. In addition, the average power consumed by the chip is approximately 26.65 µW. This is the minimal required power for the chip operating without backscatter modulation.

C. Correction of the Input Impedance

The above computed values are approximate and need to be corrected, because the voltage at pins hf_1 and hf_2 (V_0) is assumed constant, which, as we showed above, does change with the progressive power up of the IC. In order to take this phenomenon into account, the test bench in Fig. 5 is modified to that in Fig. 7.

Now the antenna is included with R_a = 10.7Ω and L_a = 64nH corresponding to Z_a = 10.7 + j363.5 at the operating frequency of 900 MHz. This time V_{hf} is assigned a value given by $(8R_aP_{av})^{0.5}$, where P_{av} should be larger than the minimal required power of 26.65 µW in order to compensate the mismatch between the antenna and the IC.

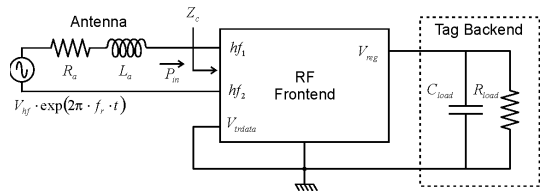


Fig.7. Test bench to correct the IC's input impedance

This simulation is closer to the actual operating state of the chip, in which the input power remains steady and the interaction between the chip and the antenna is included.

The simulation is performed once more rendering a more accurate value of Z_c. The correction process can be iterated and the value of Z_c further refined.

IV. Simulating Modulation and Demodulation

Now that the IC's input impedance in the power matching point has been determined, the required electrical parameters of the antenna (inductance and radiation resistance) are easily determined, which in turn flow into the antenna geometry during its electromagnetic design. Up to this point, the main factors affecting the operating conditions have been determined: the required input power P_{av} (the power that the

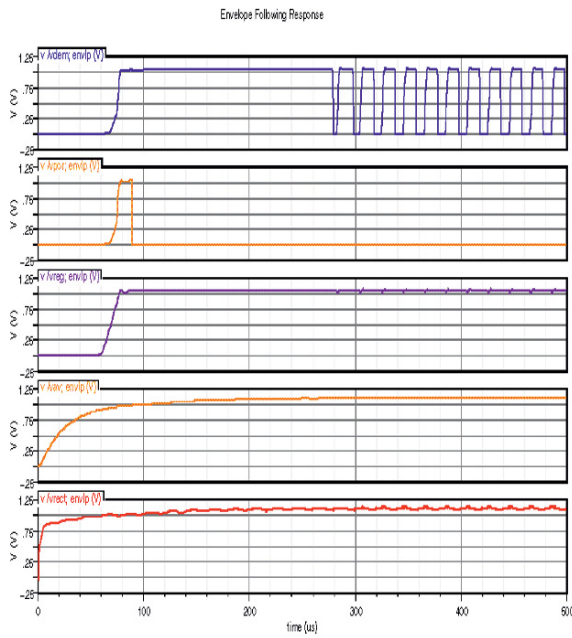


Fig. 8. Simulation of ISO18000-6 Type A demodulation

antenna should deliver to the IC) and the antenna electrical parameters (its impedance). They are independent of the chip state and should remain constant during the following simulations.

The standard **ISO/IEC 18000-6** specifies two demodulation schemes, Type A and Type B. To conform to this standard, the transponder must support at least one of these types. In this work, the transponder was designed to support both of them. Therefore, there are at least three different demodulation situations (one for Type A, two for Type B) which need to be accounted for. The test bench for the demodulator remains the same as that shown in Fig. 7. The input carrier is amplitude modulated in three different ways. Here, as an example, the simulation results for Type A are presented (Fig.8).

Simulation shows that the power up time is about 70µs. At this instant, the output voltage V_{reg} reaches its operation value of 1V and the power-on-reset (V_{por}) signal is asserted. The demodulation signal (V_{dem}) is valid after the average reference value (V_{av}) becomes stable. V_{av} and V_{rect} are the inputs to the demodulator (Fig. 9). V_{av} reaches its steady state after approximately 270us, which is within specification after the standard **ISO/IEC 18000-6**.

In order to simulate the backscatter modulation, the signal V_{trdata} in Fig. 7 is connected to a periodic pulse voltage source, which is used to simulate the backscatter control signal from the digital core. The results are shown in Fig.10, in which V_{high} and V_{low} are the DC outputs of the RF rectifier. They are used to generate the regulated voltage V_{reg}. ***V(hf2/hf1)*** is the input voltage amplitude of the chip. ***Io/hf2*** is the current into port hf_2. The backscatter power P_{bs} can be estimated using Eq.(10) [5].

$$P_{bs} = 0.5(I_1 - I_2)^2 R_a \qquad (10)$$

where I_1 and I_2 represent the current amplitudes through the antenna for the matched und unmatched cases, respectively.

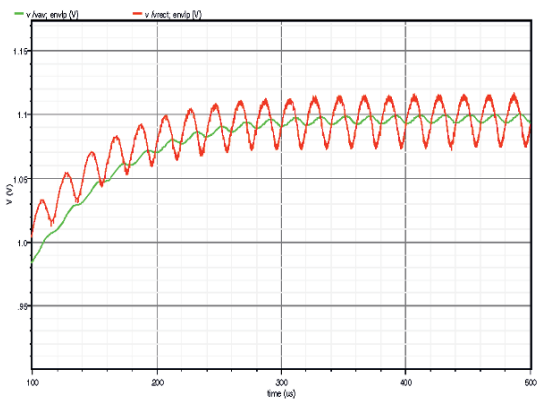


Fig.9. Simulation of V_{av} and V_{rect}

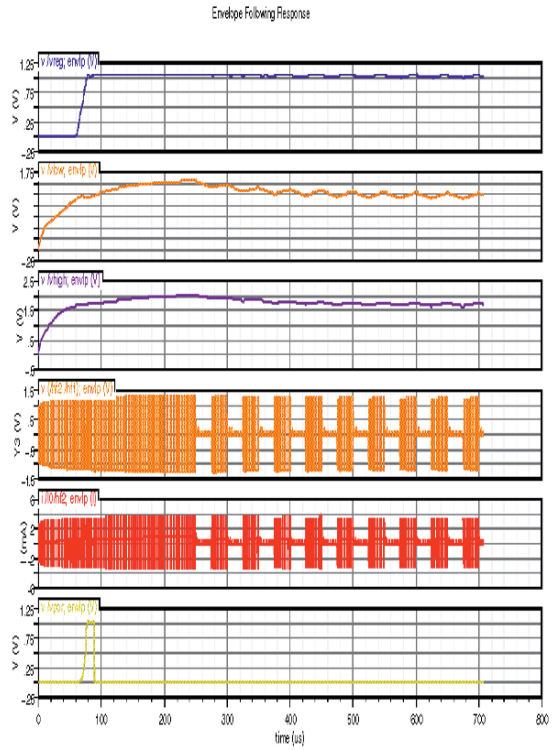


Fig.10. Simulation of the backscatter modulation

Besides the simulations shown in the present paper, other simulations such as checking if the chip can operate correctly at a very close read range, where a voltage limiter must clamp the DC output voltage, should also be performed.

V. Conclusions

In this paper, simulation issues regarding a passive UHF analog front end for an RFID transponder based on the standard ISO18000-6 are discussed. Special attention is given to the determination of the power matching point and the antenna parameters at such point. Finally, the test bench for the simulation is established, upon which several types of simulations can be performed. Based on these results, the behavior of the whole chip becomes clear before it is actually manufactured. The design parameters of the matched antenna can also be given before the measurement of the chip takes place. In this way, the design cycle of the transponder can be considerably shortened.

References

[1] www.**rfidchina**.org/
[2] www.**rfid**forum.com/
[3] Giuseppe De Vita and Giuseppe Iannaccone, “Ultra low power RF section of a passive microwave RFID transponder in 0.35 µm

BiCMOS", *IEEE MTT-S Int. Microwave Symp. Circuits Systems (ISCAS 2005)*, Kobe, Japan, May 2005, pp. 5075-5078.

[4] Giuseppe De Vita and Giuseppe Iannaccone, "Design criteria for the RF section of UHF and microwave passive RFID transponders", *IEEE Transactions on Microwave Theory and Techniques*, Vol. 53, No. 9, September 2005, pp. 2978-2990.

[5] Udo Karthaus and Martin Fischer, "Fully integrated passive UHF RFID transponder IC with 16.7- µW minimum RF input power", *IEEE Journal of Solid-state Circuits*, Vol. 38, No.10, October 2003, pp.1602-1608.

[6] T.Umeda, H. Yoshida, S. Sekine, Y. Fujita, T. Suzuki, and S. Otaka, " A 950- MHz Recifier Circuit for Sensor Network Tags with 10-m Distance," *IEEE Journal of Solid-State Circuits*, vol.41, no.1, pp.35-41,Jan.2006.

[7] Collin, R.E., "Limitations of the Thevenin and Norton Equivalent Circuits for a receiving antenna", *IEEE Antennas and propagation magazine*, vol.45, no.2, pp.119-124, April 2003.

[8] Pavel V.Nikitin and K.V.S.Rao, Intermec Technologies Corporation, http://www.ee.washington.edu/people/faculty/nikitin_pavel/papers/AMTA_2005.pdf.

Various Methods of Economical Load Distribution in Power Plant Units in Comparison to Neural Networks Method

Mohammad Taghi Ameli[#1], Saeid Moslehpour[#2], Rahmatollah Rameshpour[#3]
[#1#3], Department of Electrical Engineering, Water and Power University [#2]Department of Electrical and Computer Engineering, University of Hartford, West Hartford, CT, USA, [#3]
[#1] ameli@pwit.ac.ir
[#2] moslehpou@hartford.edu
[#3] mtameli@yahoo.com

Abstract— **Initial capital investment for setting up an electrical power plant is huge, and the annual operating costs of major and preventive maintenance, administration, labor, fuel consumption, etc., amount to a very large sum. Therefore, optimized operation of the power plants and accomplishment of the highest level of utilization is of utmost importance.**

This paper reviews the use of neural network techniques for optimal load-distribution in a network of power plant units, and using a software program. The computed results for a one-day operation of a power plant are mathematically analyzed and compared. Fuel cost is calculated by the program and compared with the Figures obtained via manual methods.

Index Terms— Economical Load Distribution, Lagrange Multiplier, Gradient, Sharing Coefficient, Neural networks

I. Introduction

The objective of optimal load distribution is to determine the share of various power generating units in supplying the total consumption load of the system under specified constraints and condition. Minimization of fuel cost is a most fundamental component of this load-distribution optimization that may also include consideration of production units' limitations, minimization of electrical transmission losses, and environmental pollution and hazards. Like fuel costs, these limitations can be formulated as a function of the quantity of generated energy. So far, several techniques are applied for optimization of load-distribution in power systems, including recursive λ method, first and second order gradient method, sharing coefficient [1].

II. Various Methods of Optimal Load-Distribution

A. Lagrangian Method

Objective Function: (Lagrange Multiplier × Constraint function + Objective function)

$$L = F_i(P_i) + \lambda \times \phi \tag{1}$$

Step 1: First, select the (desired) generating potentials of the units within the set limits of each generating unit, and second, ensure that the sum of generating potentials is equal to the total load of the system.

Step 2: Calculate the incremental losses and the total losses (PL), assuming invariant thereafter.

Step 3: Calculate Φ that makes the sum of P1, P2, … PN equal to total load plus losses. This step comprises the solution of equations.

Step 4: Compare the computed potentials with the previous iteration and if the difference is negligible, stop (otherwise, go back to step 1 and use the computed generating potentials in the next iteration).

B. Gradient Method

The gradient method is directional search computation. It always begins with a feasible solution and, in searching for the final and optimal solution; it only proceeds via the feasible paths. This means that if computation is discontinued at any stage, the solution obtained so far may be adopted as a feasible one [2].

C. Base-load Point and Sharing Coefficients Method

In the above-mentioned methods, searching for an optimal load-distribution solution consists of computations based on the available generating units and the required quantity of total load for the same time. The method of base-load point and sharing coefficients presumes that at any specific hour of the day, the system operates at a specific and optimal base point.

Considering that base point, if the consumption load varies, the outputs of generating units are accordingly moved from their previous optimal point to a new optimal point.

D. Neural Network Method

Neural networks belong to the class of intelligent dynamic systems that operate on empirical or experimental data and transfer the discovered underlying relations to the neural network structure. Neural networks are called intelligent because they exhibit learning and adaptation on the basis of configurative or numerical processing and analyses.

Neural networks are inspired by the known principles of operation of the brain and nervous system in reception and transmission of messages. The similarity of artificial neural networks and the brain and nervous system has greatly empowered the neural networks in both structural and technological developments [3].

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 226–230.

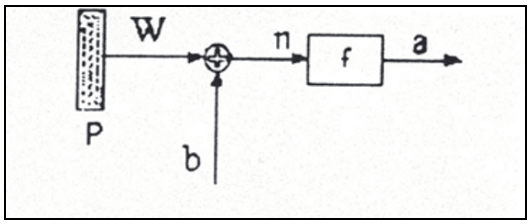


Fig.1. Single Neuron Neural Network Model
P: input vector, a: output vector, W: weight vector, b: bias vector, n: net input, f: transfer function

III. The Software Program

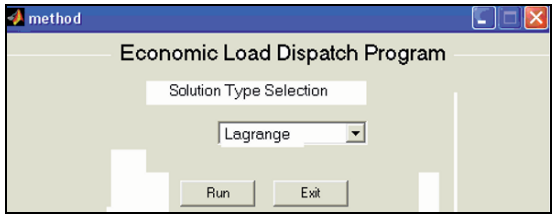


Fig.2. Window for Solution Type Selection

The starting window has a combo box for selection of the desired solution type and two keys to run or exit the program.

A. Selection of Lagrange Solution Type

Selection of Lagrange Solution Type opens the respective window for inputting the necessary data:

1. 0, a1, a2: Parameters for units' calculation of H (MBtu/h)
2. Pmin, Pmax : Minimum and maximum potency of each unit
3. Loss matrix (if required)

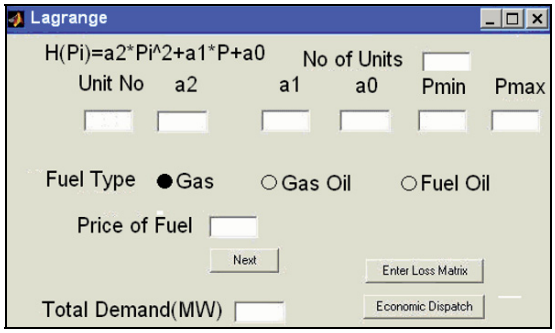


Fig3. User Interface for Lagrange Method

Upon entry of the above data, the program starts execution and outputs the optimal solution. By selection of the fuel type (natural gas, gasoil and fuel oil) and price, the program computes the fuel cost function.

B. 3-2 Selection of Gradient Method

In the window for solution type selection, "Gradient" invokes the module for this method, and a window opens for inputting the unit that should be used as the reference unit, and the parameters a2, a1, a0, Pmin and Pmax for each unit.

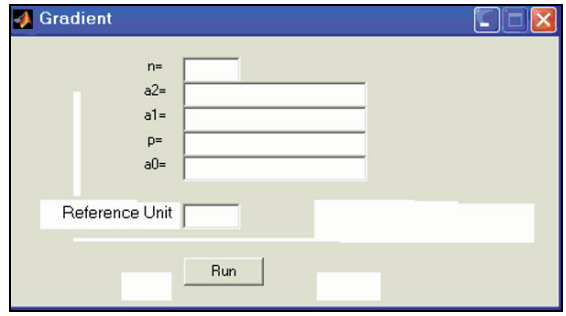


Fig. (4) – User Interface for Gradient Method

Then by pressing the Run button, the program starts execution and outputs the results.

C. Selection of Common Base Method

By selecting the "Common Base Method" in the main window, the following window is displayed for inputting:

1. Parameters of each unit
2. P: The economical potency of the unit before variation
3. dP: Change in load quantity for the unit

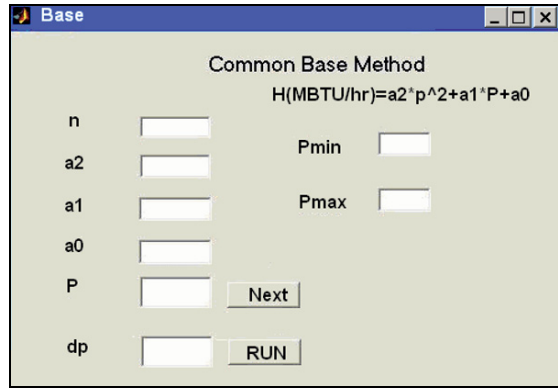


Fig. (5) – User Interface for Common Base Method

Specifying the Pmin and Pmax keeps the solution for each unit within its specified minimum and maximum limits.

D. Selection of Neural Network Method

Selection of the Neural Network option in the main window opens the respective window. On the basis of the well-know RBF and MLP methods, two programs accept the potency of

n generating inputs for 24 hour operation in the form of an n×24 matrix, perform the computations, and produce the results [4].

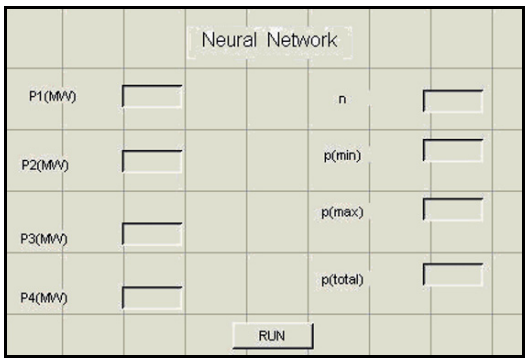


Fig. (6) – User Interface for Neural Network
N: number of units, Pmin, Pmax : minimum and maximum potency of unit, Ptotal: Total load

IV. 4- Numerical Studies

Following is a review of an example of optimal load distribution. In a 24-hour generating plant load, the optimal distribution is performed using a variety of programs and the results are compared with the Fig. that were used prior to these calculations. The differences in the costs clearly indicate the necessity of close and accurate monitoring of the load distribution [5].

For the purposes of this paper the example power plant was selected and its data was derived from the Bureau of Statistics and Production of Generation company. This power plant has four steam units with the nominal rating of 1000 MW. The fuel of this plant is natural gas and fuel oil, for summer and winter, respectively [6].

Pmin and Pmax parameters of all units of this plant are 125 and 255 respectively.

A, B, and C Coefficients:

$$H_1 = 0.0128P_1^2 + 4.909P_1 + 367.244$$
$$H_2 = 0.0053P_2^2 + 6.202P_2 + 352.628 \quad (2)$$
$$H_1 = 0.0128P_1^2 + 4.909P_1 + 367.244$$
$$H_2 = 0.0053P_2^2 + 6.202P_2 + 352.628$$

V. Manually Computed Values, Prior to Usage of This Program

The source of the following load-distribution data and values for the example power plant (Table 1) is the System Studies department of the National Network Dispatching Authority.

VI. 6- Technical and Economical Comparison of Results

Following is a single unit comparison of the optimally generated power (obtained from the software program) versus the actually generated power (on the basis of manually calculated values). In the sequel, the optimal versus actual fuel costs are compared [7].

TABLE I
ACUAL AND ECONOMIC (NEURAL MLP) GENERATION OF EXAMPLE POWER PLANT ON DEC 27.04

T.L.	4		3		2		1		Unit
	E.G.	A.G.	E.G.	A.G.	E.G.	A.G.	E.G.	A.G.	Hour
511	/3 130	/7 127	/1 125	/7 127	/3 130	/7 127	/1 125	/7 127	1
507	/7 128	/7 126	125	/7 126	/7 128	/7 126	/0 126	/7 126	2
505	/4 127	/2 126	125	/2 126	/4 127	/2 126	/0 125	/2 126	3
508	/2 129	127	125	127	/2 129	127	/0 125	127	4
509	/7 129	/2 127	125	/2 127	/6 129	/2 127	/0 125	/2 127	5
619	/9 157	/7 154	/6 151	/7 154	/9 157	/7 154	/6 151	/7 154	6
844	/3 215	211	/7 206	211	/3 215	211	/7 206	211	7
1005	/0 255	/2 251	/6 247	/2 251	/0 255	/2 251	/6 247	/2 251	8
1007	/0 255	/7 251	/0 248	/7 251	/0 255	/7 251	/0 248	/7 251	9
1000	255	250	/6 245	250	/0 255	250	/6 245	250	10
1000	255	250	/6 245	250	/0 255	250	/6 245	250	11
1000	255	250	/6 245	250	/0 255	251	/6 245	250	12
1004	255	/0 251	/3 247	251	/0 255	/5 251	/3 247	251	13
1005	255	/5 251	/8 247	/5 251	/0 255	251	/8 247	/5 251	14
1004	255	/0 251	/3 247	251	/0 255	/2 251	/3 247	251	15
1005	255	/2 251	/6 247	/2 251	/0 255	/2 251	/6 247	/2 251	16
1005	255	/2 251	/6 247	/2 251	/0 255	/2 251	/6 247	/2 251	17
1005	255	/2 251	/6 247	/2 251	/0 255	/2 251	/6 247	/2 251	18
1004	255	251	/3 247	251	/0 255	251	/3 247	251	19
1006	255	/5 251	/8 247	/5 251	/0 255	/5 251	/8 247	/5 251	20
1005	255	/2 251	/6 247	/2 251	/0 255	/2 251	/6 247	/2 251	21
1007	/0 255	/7 251	/0 248	/7 251	/0 255	/7 251	/0 248	/7 251	22
1007	/0 255	/7 251	/0 248	/7 251	/0 255	/7 251	/0 248	/7 251	23
937	239	/2 234	/5 229	/2 234	239	/2 234	/5 229	/2 234	24

E.G.: Economic Generation
A. G.: Actual Generation
T.L.: Total Load

TABLE 2
ACUAL AND ECONOMIC (NEURAL RBF) GENERATION OF EXAMPLE POWER PLANT ON DECEMBER 27, 2004

T.L.	4		3		2		1		Unit
	E.G.	A.G.	E.G.	A.G.	E.G.	A.G.	E.G.	A.G.	Hour
511	/4 130	/7 127	/5 126	/7 127	/5 130	/7 127	/4 126	/7 127	1
507	/1 130	/7 126	/3 126	/7 126	/0 130	/7 126	/1 126	/7 126	2
505	/9 129	/2 126	/2 126	/2 126	/9 129	/2 126	/0 126	/2 126	3
508	/2 130	127	/3 126	127	/2 130	127	/2 126	127	4
509	/3 130	/2 127	/4 126	/2 127	/3 130	/2 127	/2 126	/2 127	5
619	/5 149	/7 154	/9 141	/7 154	/5 149	/7 154	/0 143	/7 154	6
844	/4 219	211	/5 208	211	/4 219	211	/1 210	211	7
1005	/7 254	/2 251	/0 247	/2 251	/7 254	/2 251	/0 247	/2 251	8
1007	/0 255	/7 251	/9 247	/7 251	/0 255	/7 251	/3 247	/7 251	9
1000	/1 254	250	/3 246	250	/1 254	250	/3 246	250	10
1000	/1 254	250	/3 246	250	/1 254	250	/3 246	250	11
1000	/1 254	251	/3 246	250	/0 254	251	/3 246	250	12
1004	/6 254	/5 251	/0 246	251	/6 254	/5 251	/9 246	251	13
1005	/9 254	251	/2 247	/5 251	/8 254	251	/2 247	/5 251	14
1004	/6 254	/2 251	/9 246	251	/6 254	/2 251	/9 246	251	15
1005	/7 254	/2 251	/0 247	/2 251	/7 254	/2 251	/0 247	/2 251	16
1005	/7 254	/2 251	/0 247	/2 251	/7 254	/2 251	/0 247	/2 251	17
1005	/7 254	/2 251	/0 247	/2 251	/7 254	/2 251	/0 247	/2 251	18
1004	/6 254	251	/9 246	251	/6 254	251	/9 246	251	19
1006	/9 254	/5 251	/2 247	/5 251	/9 254	/5 251	/2 247	/5 251	20
1005	/7 254	/2 251	/0 247	/2 251	/7 254	/2 251	/0 247	/2 251	21
1007	/0 255	/7 251	/4 247	/7 251	/0 255	/7 251	/3 247	/7 251	22
1007	/0 255	/7 251	/4 247	/7 251	/0 255	/7 251	/3 247	/7 251	23
937	/2 243	/2 234	/5 233	/2 234	/2 243	/2 234	/5 234	/2 234	24

As it is shown in table 2 and fig. 7 we can conclude that our proposed method reduces fuel cost.

Result from neural networks indicates less economical in certain hours compare to Lagrange method, this is due to lack of optimization of neural networks. By increasing the number of inputs and outputs to neural networks we can decrease the error factor resulting in saving fuel.

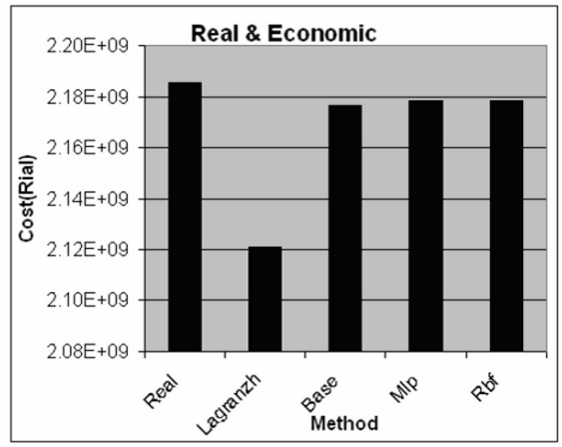


Fig. (7)- Comparison of Actual (Real) Fuel Cost vs. Computed Cost by Various Methods

VII. CONCLUSION

This paper discusses economics of load distribution amongst heating units. Achieving this goal and reduction in cost requires a smart management system.

This software can easily produce the hourly program of load distribution for various generating units. The various implemented methods of Lagrangian multiplier, Gradient, Common Base, Sharing Index and Neural Network are accurate, reliable, fast, and free from intricate complexities, therefore quite suitable as software solutions for optimal load distribution problems.

The neural network method and the fast numerical algorithms that were used yielded a better performance in optimization and load distribution of thermal power plants. The characteristic independence of our neural network method from the number of units and the network structure could circumvent the difficulties of non-linearity. Considering the simulation results, the speed and accuracy of results highlights the advantages of neural network.

The load distribution results obtained for the example thermal power plant and computation of its daily fuel costs indicate that the Lagrangian method offers a significant reduction in fuel costs.

For an optimal load distribution, the neural network method seems to be superior as its functionality is independent of the number of units and the network structure. Besides, for non-linear transfer functions that cannot be solved by mathematical methods, a neural network is desirable and can produce better results.

From our results we can conclude as long as there is adequate training of neural networks, achieving similar results however is not easily accomplished using other methods.

REFERENCES

[1] Wolenberg, Wood, "Generation, Operation and Control in Power Systems", Translated by Dr. Seifi, Tarbiat Modarress University, Tehran 1992.

[2] LittleField, Hanselman, "MATLAB UserGuide", translated by Behroozi, Tehran 2002.

[3] Farrokhi M, Jalilzadeh S, "Economical Load Distribution for Thermal Plants using Fuzzy-Neural Networks," Electrical Engineering Seminar, Water and Power University, Tehran 2004.

[4] Rastegar and Firoozian, "Optimal Load Distribution for Thermal Plants Using Fuzzy-Neural Decision-making Algorithm," Seminar of Electrical Engineering in Amir Kabir University, Tehran, Iran, 2005.

[5] T. Yalcinoz and M. J. Short, "Neural Networks Approach for Solving Economic Dispatch Problem with Transmission Capacity Constraints". IEEE Transactions on Power Systems, Vol. 13, No. 2, May 1998.

[6] A.H. Gee and R.W. Prager, "Polyhedral Combinatorics and Neural Networks", *Neural Computation*, Vol. 6, pp. 161-180, 1994.

[7] C.Y. Maa and M.A. Shanblatt, "Linear and Quadratic Programming Neural Network Analysis," *IEEE Trans. On Neural Networks*, Vol.3, No. 4, pp. 580-594, 1992.

Automated Surveillance of Intruders at US Borders

Kalyan Marneni & Sreela Sasi
Gannon University
Department of Computer and Information Sciences
Erie, Pennsylvania, USA

{marneni001, sasi001}@gannon.edu

Abstract:

There is a growing need to block the illegal immigrants into US-Mexico border as part of 'war on drugs' and for 'national security', especially after September 11, 2001 [1]. In this research, an efficient border surveillance system using video cameras and video processing units connected to a base station using a wireless network is proposed. A novel image enhancement technique is incorporated to get a quality image of the intruder even at low lighting conditions or during night. This system will send an interrupt signal to the nearby base station when an intruder is detected along with intruder's image for immediate law enforcement activities. A light source is also provided with automatic 'ON/OFF' capability to capture the image when it is completely dark.

I. INTRODUCTION

There is a growing need to block the illegal immigrants into US-Mexico border as part of 'war on drugs' and for 'national security' especially after September 11, 2001. United States has 7500-mile land and air border shared with Canada and Mexico out of which US shares 1,951 mile border with Mexico. Each year, more than 500 million people are admitted into the United States of which 330 million are noncitizens mostly from Mexico [1]. On October 26th 2006, President Bush signed "secure fence act 2006" authorizing 700 miles of fencing along the US-Mexico border, and allotted $34.8 billion [2]. Video surveillance is a powerful tool for monitoring public safety and security round the clock. The demand for video installations is growing rapidly with increasing need for more security at airports, military bases, embassies, and other critical environments.

Border monitoring has become a main issue nowadays as part of 'war on drugs' and for blocking the intruders crossing the border into the US illegally. The present system to detect the intruders is only at suspicious border areas using sensors and radio frequency. It is not only expensive to implement through out the border but also inefficient due to its reliability [2]. Hence the situation demands a reliable and an economical system that can continuously monitor the entire border area for enforcing security. Building just a fence may not be sufficient to prevent people from crossing the border, but it must be monitored round the clock. A possible video surveillance system setup to monitor the people crossing the US-Mexico border is shown in Fig 1.

Video cameras along with video processing units (VPU) are placed at a distance from the border on a pole facing the fence. VPU also consist of a light source

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 231–236.

that can automatically be turned 'ON/OFF' based on the quality of the video image. A battery will be provided in case of power failure. The number of cameras needed per

Fig. 1. Possible video surveillance system

mile is based on the focal length of the camera lens and the size of the detector for standard camera. The ground distance between the camera poles can be computed using [3] as shown in (1).

$$\frac{\text{Altitude}}{\text{Ground distance}} = \frac{\text{focal length}}{\text{size of the detector}} \text{----- (1)}$$

A number of video units are connected to a base station where the law enforcement people will be alerted with the actual location and image of the intruder through a wireless network in case anybody trespassing the border fence. Fig 2 shows a 'Monitoring Terminal' at the base station that receives video frames and alert signals from six adjacent video camera (VC) units.

Clarity of the images of intruders from video cameras during the day time will be good compared to the one at night times. In this paper, a novel approach is proposed to detect intruders not only during the day time but also during night time. A background subtraction technique that fuses the properties of 'motion of objects' in image pixels and the 'changes occurring between the adjacent input video frames' is proposed. The intensity and

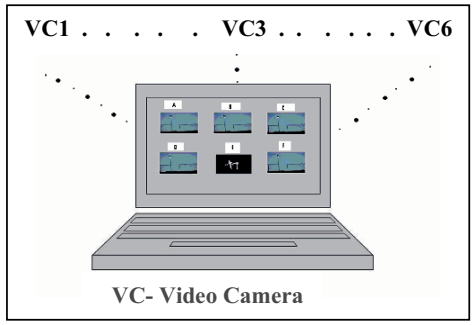


Fig. 2. 'Monitoring Terminal' at the Base Station

contrast of an image are enhanced before background subtraction to track the intruders at lower illumination levels.

II ASIB ARCHITECTURE

The processes involved in the proposed "Automated Surveillance of Intruders at the US Borders (ASIB)" system are shown in Fig 3.

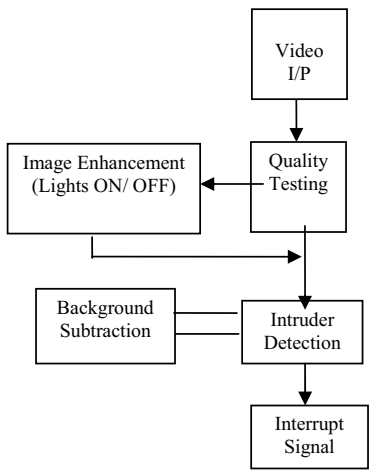


Fig. 3. Processes in the ASIB System

The images from a video camera are tested for quality, and image enhancement is done for both intensity and contrast for low quality images. These images are used

for intruder detection after background subtraction. An interrupt signal is generated if an intruder is present in the image.

A. Quality testing

In this paper, the quality of images in a video sequence is determined using image histogram. An intelligent decision is made based on various threshold values applied to the range of image histogram. A pseudo code is given below for testing the quality of the images.

```
Extract image frames from the video.
For i = 1: n (last image frame of the video)
{
 Do
Convert RGB image frame to Gray
 Scale image frame
Function Imagehistiogram ( )
For imhist 100 to 150
   {
Generate interrupt (switch on the light)
Go to Function Image enhancement ( )
For imhist (75 to 100) && (150 to 175)
        {
 Go to Function Image enhancement ( )
   else
Go to Function Background subtraction ( )
      }
    }
}
```

The threshold values are given for local geometrical conditions. If the quality of an image is determined to be good by the proposed algorithm then the image is directly sent for background subtraction. If the quality of image is determined to be poor (75-100) && (150-175) then the image is enhanced. If the quality of an image is too poor (100-150) then an interrupt is generated to switch 'ON' the light connected to the VPU and then the images are captured.

B. Image enhancement

The enhancement process for poor quality images are shown in Fig 4.

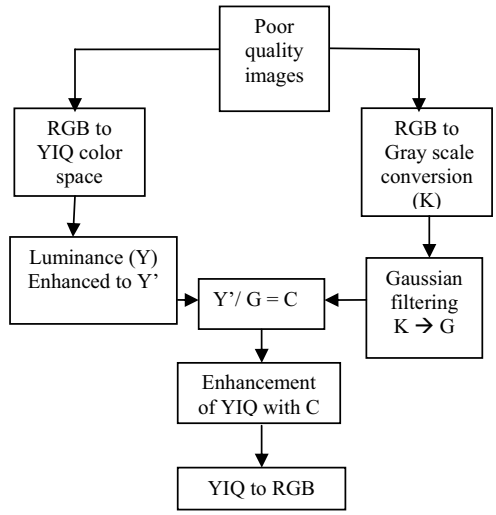


Fig. 4. Processes for image enhancement

For enhancing the images of video sequence the RGB image is first converted into YIQ color space (NTSC). In the YIQ color space, the luminance is the grayscale signal intensity used to display pictures on monochrome (black and white) televisions. The other components carry the hue and saturation information. Luminance [Y] component of an image is extracted which takes lesser time than using La*b*, and the nonlinear equation in (2) is applied for luminance enhancement [4].

$$Y_n' = \frac{Y_n^{0.24} + (1-Y_n) * 0.5 + Y_n^2}{2} \quad \text{----- (2)}$$

where Y_n being the initial luminance value and Y_n' being the enhanced luminance value. Simultaneously, the RGB image is converted into grayscale image [K], and Gaussian filtering is applied to reduce the noise. The Gaussian smoothing operator is a 2-D convolution operator that is used to `blur' images and remove noise. In 2-D, an isotropic (*i.e.* circularly symmetric) Gaussian has the form in (3):

$$G(x) = \frac{1}{\sqrt{2\Pi\,\sigma}} e^{-\frac{x^2}{2\sigma^2}\theta} \quad \text{---------- (3)}$$

A constant 'C' is obtained by dividing the enhanced luminance, 'Y_n'' of the image by gaussian filtered gray scale image, 'G'. This constant 'C' is used for enhancing the luminance and chrominance components of YIQ color space. Then these enhanced components are combined to obtain the enhanced RGB image.

A poor quality frame from a video taken near a fence at a parking lot for simulation purpose is shown in Fig 5 along with the corresponding enhanced image.

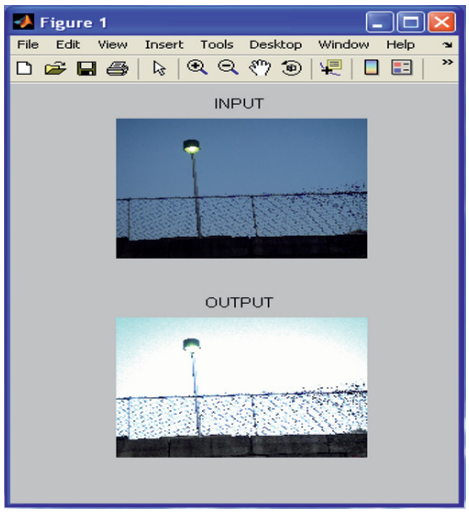


Fig. 5: Enhanced image under low illuminations

C. Intruder Detection

The most common approaches to identify the moving objects are optical flow [5] and background subtraction [6], [7] based methods. Optical flow is used to detect independently moving objects in the presence of camera motion. However, most optical flow based methods are computationally complex and cannot be applied to full-frame video streams in real-time without specialized hardware. Though there are many background subtraction techniques [6], [7] available background updating for complex background scenarios has been challenging. The important tasks of background subtraction method used in this paper are background modeling and foreground detection.

1) Background Modeling

Background modeling is at the heart of any background subtraction algorithm. In this paper a modified version of Median filtering [8] is used for background estimation. A buffer of previous 'L' video frames are stored, and estimates the background image based on the temporal variation of each pixel within the buffer as shown in (4). Non-recursive techniques are highly adaptive as they do not depend on the history beyond those frames stored in the buffer.

2) Foreground detection

The input video frame is compared with the background model, and identifies candidate foreground pixels for foreground detection. The most commonly used approach for foreground detection is to check whether the input pixel is significantly different from the corresponding background estimate.

$$\frac{Ft(x, y) - Bt(x, y) - \mu_d}{\sigma_d} > T_s \text{ -------- (4)}$$

where μ_d and σ_d are the mean and the standard deviations of Ft(x, y) - Bt(x, y) for all spatial locations (x, y), and T_s is the threshold estimate for the background model. The detection of an intruder from a background subtracted image is done as follows:

```
For i = 1: n
{
For j = 1: x
 {
For k = 1: y
    {
Read pixel color
If (pixel = = white)
        {
Count = count+1;
         }
     }
 }
Compare count with respect to threshold height (5ft is used currently)
If (Count above threshold height) > (2/3 of count below threshold height)
  {
 Image counter = Image counter +5
Generate interrupt signal
  }
}
```

An '*INTERRUPT*' signal is sent to the base station wirelessly, which will generate an 'alarm signal', followed by the video frames containing the intruder's image. This will not only help the law enforcement people to locate the position, but also helps with the intruder's image for tracking him down at the earliest. In order to avoid false alarm generations, a height corresponding to the fence level is set for each image. So, an alarm signal is generated if and only if an intruder's image is incrementally increasing in adjacent video frames.

III. SIMULATION

Simulation is conducted on video images taken under nonuniform and insufficient lighting conditions near a fence at a parking lot. The fence is 12 feet high. A threshold height in the image is marked corresponding to an actual height of 5ft of the fence. If the number of white pixels above the preset height (5 ft) is 2/3 of total pixels, then the intruder will be traced. Few of the frames containing intruder is shown in Fig. 6-10. The accuracy of detection of intruders is given in Table 1.

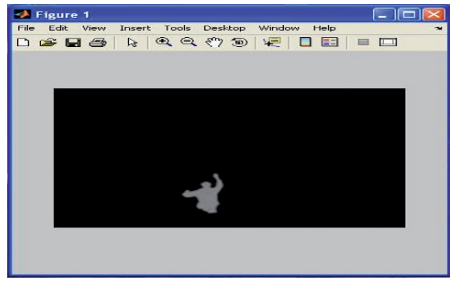


Fig. 6. Frame 34

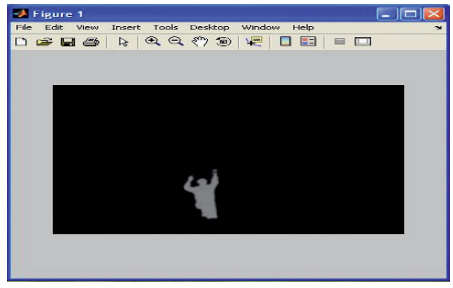


Fig. 7. Frame 41

In case of Figure 10, where the intruder is not fully included in the frame, there is a chance of incorrect detection as shown in the 5th row of Table 1. In order to overcome such situations the ground distance between

adjacent poles must be less by a foot than the distance computed using equation (1). Hence the complete fence will be covered by at least one of the cameras.

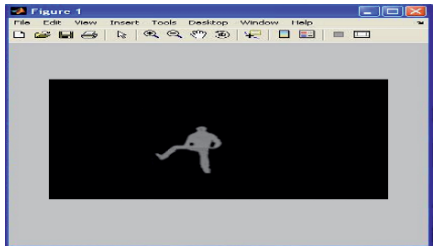


Fig. 8. Frame 69

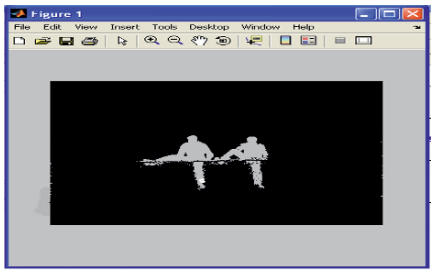


Fig. 9. Frame 104

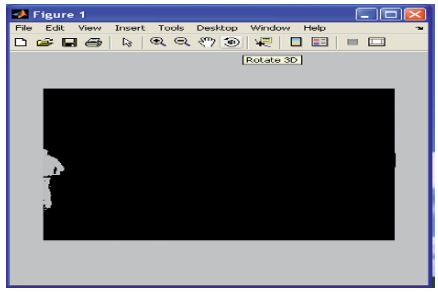


Fig. 10. Frame 164

IV. CONCLUSION

ASIB system is a simple video surveillance system. Hence it is economical and efficient as the installation cost and power consumption is less. This set up can be used not only at the national borders but also at military perimeters and embassies. Extensive testing is needed near the US-Mexico border to fine tune the ground distance between adjacent poles, and to determine the positioning of pole from the fence The future work would be to modify the algorithm to support the infrared cameras if it becomes cheaper.

Table 1. Accuracy of Detection

S.no	Figures	Accuracy of detection (%)
1	6	0
2	7	100
3	8	100
4	9	100
5	10	0

References:

[1] www.whitehouse.gov/homeland/bordersecurity
[2]http://www.globalsecurity.org/org/news/2006/061107-border-security.htm
[3]www.nasa.gov/audience/forresearchers/features/index.html.
[4] Li Tao and K. Vijayan Asari, “An efficient illuminance-reflectance nonlinear video stream enhancement model,” *IS&T/SPIE Symposium on Electronic Imaging:* Real-Time Image Processing III, San Jose, CA, vol. 6063, pp. 60630I-1-12, January 15-19, 2006.
[5] J. Barron, D. Fleet, and S. Beauchemin, Performance of optical flow techniques”, *International Journal of Computer Vision*, 12(1): Pages: 42–77, 1994
[6] Haritaoglu, I. Harwood, D., Davis, L.S., “W4: real-time surveillance of people and their activities”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume: 22, Issue: 8, Pages: 809 – 830. Aug. 2000
[7] Pakorn KaewTrakulPong, Richard Bowden, “A real time adaptive visual surveillance system for tracking low resolution color targets in dynamically changing scenes”, *Image and Vision Computing* 21, Pages:913-929, 2003
[8] P. KawTraKulPong, and R. Bowden, “An improved adaptive background mixture model for real-time tracking with shadow detection,” *Proceedings of Second European Workshop on Advanced Video-based Surveillance Systems*, 2001

PDFF and H^∞ Controller Design for PMSM Drive

Stone Cheng†, Yuan-Yong Huang, Hsin-Hung Chou, Chia-Ming Ting, Cheng-Min Chang, Ying-Min Chen

†Department of Mechanical Engineering, National Chiao-Tung University, Hsinchu, Taiwan, ROC

Mechatronics Control Dept., Intelligent Machinery Technology Div., MSL/ITRI, Hsinchu, Taiwan, ROC

***Abstract*- This paper presents analysis, design and simulation of velocity loop PDFF controllers and H^∞ feedback controller for permanent magnetic synchronous motor (PMSM) in the AC servo system. PDFF and H^∞ control algorithm have its own capability of achieving good performance criteria such as dynamic reference tracking and load torque disturbance rejection. The PDFF is designed and analyzed in the forward loop to provide low frequency stiffness and overcome low-frequency disturbances like friction. While in the feedback loop, H^∞ controller is designed to meet system robust stability with the existence of external disturbance and model perturbations. The proposed PDFF and H^∞ controllers are designed based on the transfer function of the poly-phase synchronous machine in the synchronous reference frame at field orientation control (FOC). The parameter variations, load changes, and set-point variations of synchronous machine are taking into consideration to study the dynamic performance.**

Keywords: PMSM, H^∞ feedback control, PDFF controller,

I. Introduction

The PMSM motor servo drive play an important role in industrial motion control applications including machine tools, factory automation and robotics in the low-to-medium power range. Several situations encountered in these applications: 1) Plant parameters such as load inertia and friction may vary during operation as the payload changes. 2) System bandwidth is limited by the presence of a tensional resonance of the mechanical system. 3) In AC servo motors, higher torque ripple and coupled dynamics with magnetic flux caused the nonlinearities in torque response and torque transients. 4) The set-point tracking capability in both dynamic and steady-state conditions and the load torque disturbance rejection capability are varying during applications. Several control techniques [1-7] have been developed to overcome these issues. Derived from generalized PID controller, the PDFF controller is allowing the user to eliminate overshoot and provide much more DC stiffness than PI by properly choosing the controller parameters. It is also known [8] that PDFF controller is less sensitive to plant parameter variations and its disturbance rejection characteristics are much better than that of the PI controller. Along with PDFF controller, H^∞ control theory is one of the successful algorithms for robust control problem in PMSM drive to provide better tolerance to disturbance and modeling uncertainties. In this paper, the H^∞ design procedure[4,9,10] is proposed and consists three main stages: 1) using weighting matrices W_1 and W_2 to shape the singular values of the nominal plant follows the elementary open-loop shaping principles; 2) the normalized coprime factor H^∞ problem is used to find a robust central controller stabilizing this shaped plant, and the observer is obtained from the left coprimeness of the central controller; 3) the H parameter in the controller is used as a tradeoff between robust stability and performance.

II. Mathematical Model of the PMSM

The field orientation of the PMSM is defined as d-axis, and q-axis that leads the d-axis 90 electric degrees. In the d-q coordinates, the PMSM voltage-current and flux equations are shown as follows:

$$v_d = Ri_d + \dot{\lambda}_d - \omega_r \lambda_q \tag{1}$$

$$v_q = Ri_q + \dot{\lambda}_q - \omega_r \lambda_d \tag{2}$$

$$\lambda_d = L_d i_d + \lambda_{PM} \tag{3}$$

$$\lambda_q = L_q i_q \tag{4}$$

Where v_d and v_q are voltages of the d, q axis; R is the stator resistance; i_d and i_q are the d, q axis stator currents; ω_r is the rotor speed; λ_d and λ_q are the d, q axis flux induced by the currents of the d, q axis inductance; L_d and L_q are the q, d axis inductances with the same value, and λ_{PM} the constant mutual flux of the permanent magnet.

When the stator current vector is oriented perpendicular to the rotor magnetic field, the field-oriented control for PMSM yields i_d =0. In the case, the electromagnetic torque is in strict positive proportion to i_q:

$$T_e = \frac{3P}{4}\lambda_{PM} i_q = K_T i_q \tag{5}$$

where P is the number of poles and K_T is the motor torque constant.

The mechanic motion equation is:

$$T_e = K_T i_q = T_d + B\omega_r + J\frac{d\omega_r}{dt} \tag{6}$$

where J is the moment of inertia; B is the viscous friction, and T_d is the torque disturbance such as the load resistance, the torque ripple and the resistance caused by nonlinear factors.

III. Design of the Control System

A. Control Scheme

The proposed control scheme is presented in Fig. 1 where the nominal plant is $G(s) = 1/(Js+B)$; $K(s)$ is the velocity feedback controller designed by the loop shaping design procedure (LSDP) and the algebraic method, and the velocity lop controller is a PDFF controller. $K(s)$ is used for attenuating the disturbance T_d, and plant uncertainty, and the PDFF controller is used as velocity loop adjuster to improve the low-frequency stiffness.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 237–241.

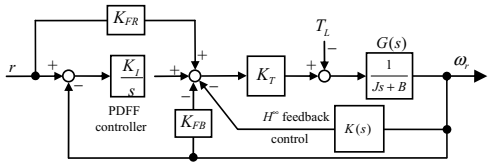


Fig. 1 Control scheme

B. Velocity Feedback Controller

In this paper, a continuous time control design approach based on H^∞-optimization control design is performed for a model of the PMSM system as seen from the digital computer control design approach. Consequently, performance is specified at the controller disturbance instants.

Minimum phase W_1 and W_2 are proper stable, real rational function denoted by RH_∞.The left and right coprime factorizations of W_1GW_2 are $\tilde{M}_S^{-1}\tilde{N}_S$ and $N_S M_S^{-1}$, respectively. Moreover, a doubly coprime factorization exists as follows:

$$\begin{bmatrix} X_r & Y_r \\ -\tilde{N}_S & \tilde{M}_S \end{bmatrix}\begin{bmatrix} M_S & -Y_l \\ N_S & X_l \end{bmatrix}=\begin{bmatrix} M_S & -Y_l \\ N_S & X_l \end{bmatrix}\begin{bmatrix} X_r & Y_r \\ -\tilde{N}_S & \tilde{M}_S \end{bmatrix}=I \tag{7}$$

where N_S, M_S, $\tilde{N}_S$, $\tilde{M}_S$, X_r, Y_r, X_l, and Y_l are over RH_∞. Then, the velocity controller $K(s)$ is defined as follows:

$$K(s)=W_1(s)K_v(s)W_2(s) \tag{8}$$

Where $K_v(s)=[X_r+H^{-1}Y_l\tilde{N}]^{-1}[Y_r-H^{-1}Y_l\tilde{M}]$ and H is a unit over RH_∞. With $K(s)$ of (8), the velocity feedback loop is internally stable. Moreover, X_r and Y_r of $K_v(s)$ in (8) play the similar role as central controller although H in $K_v(s)$ cannot be 0. According to this property, X_r and Y_r can be designed using the LSDP and H will be used to reject step and sinusoidal disturbance, as follows.

C. Design of Velocity Controller Using the LSDP and the Algebraic Method

The first stage in the LSDP uses a pre-matrix W_1 and/or a post-matrix W_2 to shape the singular values of the nominal plant G as a desired open-loop shape $G_S = W_2GW_1$. Constant or dynamic W_1 and W_2 are selected such that G_S has no hidden modes. Constant weighting matrices can improve the performance at low frequencies and increases the crossover frequency. Moreover, the dynamic W_1 or W_2 is used as the integral action with the phase-advance term for rejecting the input and output step disturbances. W_1 or W_2 is selected as the diagonal matrix and each principal element is $(s+\phi)/s$ where $\phi>0$ is lower than the crossover frequency. The integral action improves the performance at low frequencies, and the phase-advance term $s+\phi$ avoids the slope of the open-loop shaping at the crossover frequency more than -2, and adjusts the robustness in the feedback system. If ϕ is closer to the imaginary axis, the robustness is larger. The stage is the same as the velocity controller herein.

[11-14] advocate an expression of coprime factor uncertainty in terms of additive stable perturbations to coprime factors of the nominal plant. Such a class of perturbations has advantages over additive or multiplicative unstructured uncertainty model. For example, the number of unstable zeros and poles may change as the plant is perturbed. The perturbed plant [See Fig. 2.] is written

$$G_\Delta=(N_S+\Delta_N)\cdot(M_S+\Delta_M)^{-1} \tag{9}$$

where the pair (M_S, N_S) is a normalized right coprime factorization of G_S, and Δ_M and Δ_N are stable, unknown transfer functions representing the uncertainty and satisfying $\left\|\begin{bmatrix}\Delta_N \\ \Delta_M\end{bmatrix}\right\|_\infty<\varepsilon$, where $\varepsilon(>0)$ presents the stability margin.

In the second stage of the LSDP, the robust stabilization H_∞ problem is applied to the normalized right coprime factorization of G_S, and obtains a robust controller K_∞ satisfying

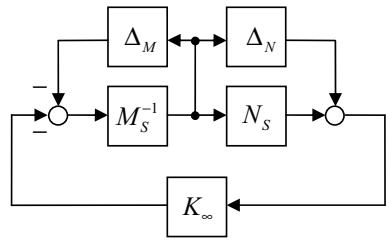


Fig. 2 Right coprime factor robust stabilization problem

$$\left\|M_S^{-1}(I+K_\infty G_S)^{-1}\begin{bmatrix}K_\infty & I\end{bmatrix}\right\|_\infty\le\varepsilon^{-1} \tag{10}$$

Suppose the shaped plant of G_S has the minimal realization (A, B, C, D). A central controller satisfying (10) is obtained as follows [15]:

$$K_\infty=\left[\begin{array}{c|c} A+BF+\gamma^2(W^T)^{-1}ZC^T(C+DF) & -\gamma^2(W^T)^{-1}ZC^T \\ \hline B^TX & D^T \end{array}\right] \tag{11}$$

where $F=-S^{-1}(D^TC+B^TX)$; $W=I+(XZ-\gamma^2I)$, and X and Z are the solutions to the two algebraic Riccati equations as follows:

$$(A-BS^{-1}D^TC)^TX+X(A-BS^{-1}D^TC)-XBS^{-1}B^TX+C^TR^{-1}C=0 \tag{12}$$

$$(A-BS^{-1}D^TC)Z+Z(A-BS^{-1}D^TC)^T-ZC^TR^{-1}CZ+BS^{-1}B^T=0 \tag{13}$$

where $R=I+DD^T$, and $S=I+D^TD$.

If the plant is assumed to be strictly proper, i.e. $D=0$, the realizations for the doubly coprime factorization can be presented as follows.

$$\begin{bmatrix} M_S \\ \hline N_S \end{bmatrix} = \left[\begin{array}{c|c} A+BF & B \\ \hline F & I \\ \hline C & 0 \end{array}\right] \tag{14}$$

$$\begin{bmatrix} \tilde{N}_S \vdots \tilde{M}_S \end{bmatrix} = \left[\begin{array}{c|c:c} A+QLC & B & QL \\ \hline C & 0 & I \end{array}\right] \tag{15}$$

$$\begin{bmatrix} X_r \vdots Y_r \end{bmatrix} = \left[\begin{array}{c|c:c} A+QLC & B & -QL \\ \hline -F & I & 0 \end{array}\right] \tag{16}$$

$$\begin{bmatrix} X_l \vdots Y_l \end{bmatrix} = \left[\begin{array}{c|c:c} A+BF & B & -QL \\ \hline C & I & 0 \end{array}\right] \tag{17}$$

The pair ($\tilde{N}_S$, $\tilde{M}_S$) in (15) is the left coprime factorization of G_S, but not the normalized left coprime factorization. Moreover, the pair (X_r,Y_r) are the left coprime factorization of K_∞ when $D = 0$. That is, $K_\infty = X_r^{-1}Y_r$. The result presents for the second stage of the velocity controller that the pair (X_r,Y_r) in $K_v(s)$ of (8) can be obtained from the left coprime factorization of K_∞ when $D = 0$.

In Fig. 1, the transfer function from T_L to ω_r is (18).

$$\omega_r = -W_1 N_S (X_r + H^{-1} Y_l \tilde{N}) W_1^{-1} \cdot T_L \tag{18}$$

For a step in T_L, ω_r with the zero steady state must satisfy the following equation, according to the final value theorem.

$$(X_r + H^{-1} Y_l \tilde{N}_S)\big|_{s=0} = (H - M_S + X_r^{-1})\big|_{s=0} = 0 \tag{19}$$

For rejecting a sinusoidal disturbance with known frequency σ in T_L, the following equation must be satisfied obviously.

$$(X_r + H^{-1} Y_l \tilde{N}_S)\big|_{s=j\sigma} = (H - M_S + X_r^{-1})\big|_{s=j\sigma} = 0 \tag{20}$$

Hence, for rejecting a step and/or sinusoidal disturbance in T_L, H can be designed algebraically. For example, if only the step disturbance exists in T_L, H is designed to be constant as follows.

$$H = (M_S - X_r^{-1})\big|_{s=0} \tag{21}$$

If only a sinusoidal disturbance with known frequency σ_1 exists in T_L, H needs two unknown coefficients and is designed as follows:

$$H(s) = h_1 \frac{s + k_1}{s + p} \tag{22}$$

where H of (22) satisfies

$$H(s)\big|_{s=j\sigma_1} = (M_S - X_r^{-1})\big|_{s=j\sigma_1} \tag{23}$$

p(>0) is given, and h_1 and k_1 can be solved according to (23). Analogously, if a number of n sinusoidal disturbances with n known frequencies σ_1~σ_n, H needs $2n$ coefficients to be solved as follows.

$$H(s) = h_1 + \frac{h_2}{s+p} + \frac{h_3}{(s+p)^2} + \cdots + \frac{h_{2n}}{(s+p)^{2n-1}} \tag{24}$$

Hence, since the pair (X_r ,Y_r) in K_v is the left coprime factorization of K_∞ in the LSDP, the completed velocity controller has several properties of the LSDP, including consideration of plant input and output performance, limited deteriorations at plant input and output, and bounded closed-loop objective functions. The three major properties of the LSDP are listed in [16]. Moreover, the velocity controller can use the H parameter to reject step and/or sinusoidal disturbances.

The velocity feedback loop also has robustness with coprime factor uncertainty, and satisfies the following robust inequality:

$$\left\| \begin{bmatrix} Y_r - H^{-1} Y_l \tilde{M}_S & X_r + H^{-1} Y_l \tilde{N}_S \end{bmatrix} \right\|_\infty \le \varepsilon_v^{-1} \tag{25}$$

where ε_v is the stability margin in the velocity feedback loop. Eq. (25) presents that the H parameter can affect the value of the stability margin ε_v. Herein, H is selected according to the control requirements and then the value of ε_v can be checked. H may need several redesigns to obtain a satisfactory value of ε_v. Moreover, for the sake of the numerical realization, K_v also can be written as $K_v = (1 + C_v X_r)^{-1} C_v Y_r$ where $C_v = H - M_S$.

D. PDFF Velocity Control Method

In digital control systems of PMSM drive, most of applications are using its velocity and torque control mode. The position loop of PMSM drive is taken control by outside multi-axis controller such as CNC controller. Many manufacturers use PI velocity loops, eliminating the derivative ("D") term. Tuning PI loop is easy and is ideal for maximum responsiveness applications such as pick-and-place machines. But PI control has a weakness—because of its integral gain must remain relatively small to avoid excessive overshoot provides that it does not have good low frequency "stiffness". PDFF velocity control was developed to combat this problem. Fig. 3 shows the block diagram in frequency domain of a plant with a PDFF controller of the form:

$$u(s) = d(s) + K_{FR} \cdot r(s) + \frac{K_I}{s} \cdot e(s) - K_{FB} \cdot y(s) \tag{26}$$

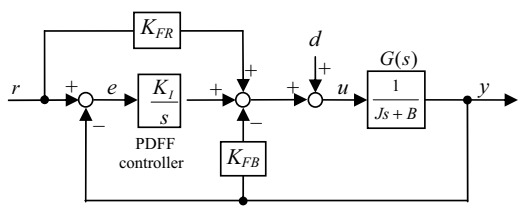


Fig. 3 Plant and disturbance with PDFF Controller

The transfer function of disturbance to output with the plant is simplified as a first order model is derived by

$$G_d(s) = \frac{s/J}{s^2 + (B/J + K_{FB}/J)s + K_I/J} = \frac{y(s)}{d(s)} \tag{27}$$

One of the most important specifications in many motion control applications is the load-torque disturbance rejection capability. The disturbance response can be tuned by moving closed poles more to the left side in the complex plane, and tracking response can be further optimized by adding zeros to

the system via feedforward, as shown in (28).

$$G_c(s)=\frac{\left(K_I+K_{FR}s\right)/J}{s^2+\left(B/J+K_{FB}/J\right)s+K_I/J}=\frac{y(s)}{r(s)} \quad (28)$$

The PDFF controller which locates the zero at an optimal place that shortens the step response rise time without overshoot.

IV. Results of simulation research

A 1KW PMSM is included in the simulation, its mechanical parameters are: J = 6.37 and B = 0.1. According to the method discussed in part C of Section III, W_1, W_2, X_r, Y_r, H and C_v are given as follows.

$$W_1=\frac{5\times10^3(s+2500)}{s},\ W_2=1,$$

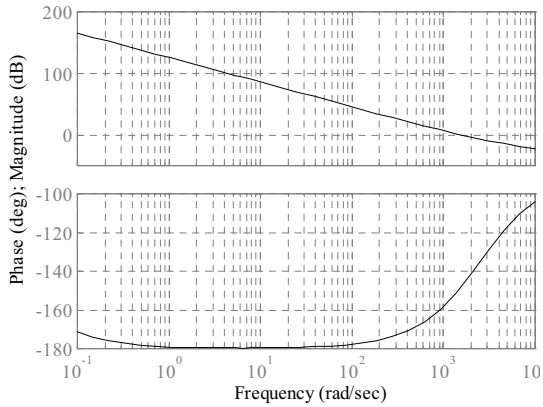
$$X_r=\frac{s^2+1.395\times10^4 s+2.348\times10^7}{s^2+1.181\times10^4 s+1.216\times10^7},$$

$$Y_r=\frac{2.016\times10^4 s+1.216\times10^7}{s^2+1.181\times10^4 s+1.216\times10^7}$$

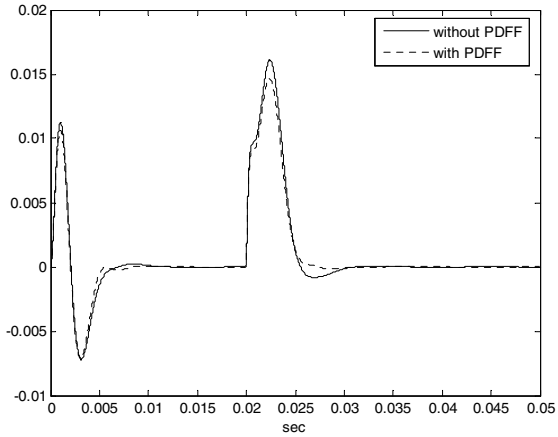
$$H=\frac{-0.393(s+2.713\times10^3)}{s+1}$$

$$C_v=\frac{-1.393s^3-1.903\times10^3 s^2-3.040\times10^6 s-2.090\times10^9}{s^3+2.132\times10^3 s^2+1.965\times10^6 s+1.962\times10^6}$$

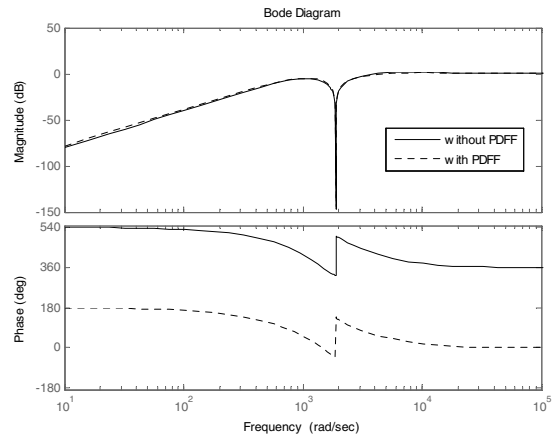
The design yields that G_S has the crossover frequency about 300Hz as shown in Fig. 4(a), and the velocity feedback loop have the stability margin 19.36%. Moreover, it yields that the velocity feedback loop can reject the 250Nm step at 0.02 sec and 300Hz sinusoidal at 0 sec disturbances in T_L as shown in Fig. 4(b), and the input sensitivity, $W_1M_SX_rW_1^{-1}$ is presented in Fig. 4(c). The effect of PDFF controller also has contribution on the disturbance rejection, as shown in Fig. 4(b).



(a)



(b)



(c)

Fig. 4 (a) G_S shape (b) disturbance responses with 250Nm step (at 0.02sec) and sin600πt (at 0 sec) (c) input sensitivity

The key difference between PI and PDFF is that PDFF forces the entire error signal through integration. This makes PDFF less responsive to the velocity command than PI. Although the feed-forward term injects the command ahead of the integral making the system more responsive to commands, moving average (MA) filter of error signal is considered to improve the responsiveness. Fig. 6 shows the step response of a 1KW PMSM and servo drive system with MA filter compensation in the velocity loop.

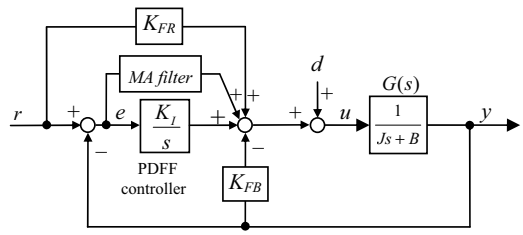
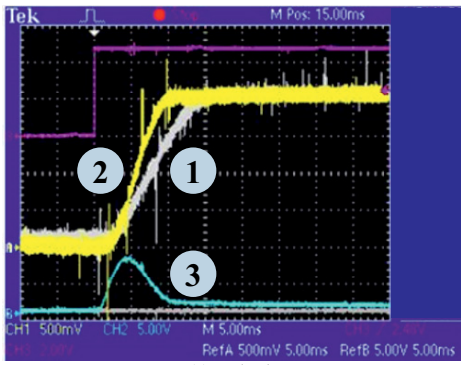
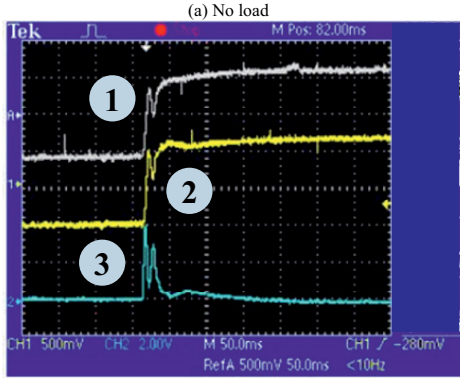


Fig. 5 Block diagram of PDFF controller with MA filter.



(a) No load



(b) With load

Fig. 6 1KW PMSM drive step response: (a) no load (b) with load. ①w/o compensation, ②with compensation, ③MA compensation signal

V. CONCLUSIONS

This paper proposes a combined design for the velocity controller of a high performance PMSM speed servo using PDFF and H^{∞} feedback control to meet the requirements of robust stability, exterior load disturbances rejection, low-frequency stiffness and responsiveness. The simulation and experimental results demonstrate the good control performance of the proposed control scheme.

ACKNOWLEDGMENT

Research supported by MSL project, ITRI, Taiwan, ROC.
Corresponding author E-mail: stonecheng@mail.nctu.edu.tw

REFERENCES

[1] T.-L. Hsien, Y.-Y. Sun, M.-C. Tsai, "*H*∞ control for a sensorless permanent-magnet synchronous drive" *IEE Proc-Electr. Power Appl., Vol. 144, No.3, May 1997,* pp. 173-181

[2] Xie Dongmei, Qu Daokui, Xu Fang, "Design of H_{∞} Feedback Controller and IP-Position Controller of PMSM Servo System" Proceedings of the IEEE International Conference on Mechatronics & Automation Niagara Falls, Canada, July 2005

[3] Jong-Sun Ko, Hyunsik Kim, and Seong-Hyun Yang, "Precision Speed Control of PMSM Using Neural Network Disturbance Observer on Forced Nominal Plant", Proceedings of the 5th Asian Control Conference, July 2004

[4] Tom Oomen, Marc van de Wal, and Okko Bosgra, "Exploiting H∞ Sampled-Data Control Theory for High-Precision Electromechanical Servo Control Design", Proceedings of the 2006 American Control Conference, June, 2006

[5] Pragasen Pillay, Ramu Krishnan "Control Characteristic and Speed Controller Design for a High Performance Permanent Magnet Synchronous Motor Drive" *IEEE Trans. Power Elec..*, vol. 5, No. 2, April 1990, pp. 151-159.

[6] S.M. Zeid, T.S. Radwan and M.A. Rahman, "Real-Time Implementation of Multiple feedback loop control for a Permanent Magnet Synchronous Motor Drive" *IEEE Proc .Canadian Conf on Elec. And Comp. End..* pp. 1265-1270, 1999.

[7] Wenhuo Zeng and Jun Hu, "Application of Intelligent PDF Control Algorithm to an Electrohydraulic Position Servo System" Proc. Of the 1999 *IEEE/ASME, Int. Conf. on Advanced Intelligent Mechatronics.* pp. 233-238.

[8] Z Nagy and A Bradshaw "Comparison of PI and PDF controls of a Manipulator ARM" *UKACC Int. Conf. on Control '98*, pp. 739-744.

[9] Yuan-Yong Huang "Robust Design of Compensator with Vidyasagar's Structure" *Ph.D Thesis, 2007, Dept. Of Mechanical Engineering, National Chiao-Tung University, Hsinchu, Taiwan, ROC*.

[10] Ali Saberi, Anton A. Stoorvogel, Peddapullaiah Sannuti, "Analysis, design, and performance limitations of H∞ optimal filtering in the presence of an additional input with known frequency", Proceedings of the 2006 American Control Conference, June, 2006

[11] M. Vidyasagar, *Control System Synthesis: A Coprime Factorization Approach*. Cambridge, MA: M.I.T. Press, 1985.

[12] M. Vidyasagar, "The graph metric for unstable plants and robustness estimates for feedback stability," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 403-417, 1984.

[13] M. Vidyasagar and H. Kumira, "Robust controllers for uncertain linear multivariable systems," *Automatica*, pp. 85-94, 1986.

[14] T. T. Georgiou and M. C. Smith, "Optimal robustness in the gap metric," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 673-686, June 1990.

[15] K. Glover and D. McFarlane, "Robust stabilization of normalized coprime factor plant descriptions with H_{∞}-bounded uncertainty," *IEEE Trans. Automat. Contr.*, vol. 34, pp. 821-830, Aug. 1989.

[16] D. McFarlane and K. Glover, "A loop shaping design procedure using H_{∞} synthesis," *IEEE Trans. Automat. Contr.*, vol. 37, no. 6, pp. 759-769, June 1992.

On Facilitating the Process of Providing Expert Advises Applying Association Rules

S. Encheva
Stord/Haugesund University College
Bjørnsonsg. 45, 5528 Haugesund
Norway

S. Tumin
IT Dept., University of Bergen
PO Box 7800, 5020 Bergen
Norway

Abstract-Relations between a trilattice and corresponding meet-distributive lattices are discussed. The three meet-distributive lattices illustrate the five information levels, five logical levels and five levels of constructivity respectively. While the trilattice shows connections among the sixteen truth values in general, the three meet-distributive lattices visualize specific information about the sixteen truth values with respect to information, logic and constructivity.

I. INTRODUCTION

Decision support systems allow significant increase of quality in the process of optimal decision selection among a large number of alternatives presented to a human-operator. The contribution of decision support systems is especially appreciated when human-operator should select important decisions, for example, in automated control systems, in real time and under conditions of absence of full prior and current information, i.e. in situations of uncertainty.

The structure of a decision support system, mathematical methods used in their design, and criteria for the optimal decisions search depend significantly on a system's purpose, type of tasks, experience of human-operator (the person making decisions), and the level of information uncertainty. In any case the analysis of various factors influencing the process of decision making, selection and registration of their interaction plays an important role.

In this paper we propose use of sixteen-valued logic for dealing with incomplete and inconsistent information. The five information levels, five logical levels and five levels of constructivity in trilattice of sixteen truth values are arranged in three meet-distributive lattices. One of the interesting observations of involves computation of '*a sentence is constructively refute* $\wedge$ *a sentence is rejectable*'. The conclusion is based on the rule that a conjunction is true if and only if both conjuncts are true. Since there is no truth value such that both *a sentence is constructively refuted* and *a sentence is rejectable* have it, the result is the empty set.

The rest of the paper is organized as follows. Related work and definitions and statements from formal concept analysis and rule mining may be found in Section II. The main results of the paper are placed in Section III and Section IV. The paper ends with a conclusion in Section V.

II. BACKGROUND

Formal concept analysis [27] started as an attempt of promoting better communication between lattice theorists and users of lattice theory. Since 1980's formal concept analysis has been growing as a research field with a broad spectrum of applications. Various applications of formal concept analysis are presented in [14].

The complexity of mining frequent item sets is exponential and algorithms for finding such sets have been developed by many authors such as [6], [11], and [28].

Mining association rules is addressed in [1] and [3]. Algorithms for fast discovery of association rules have been presented in [6], [21], and [29].

Broad decision support systems knowledge base providing generalizations and directions for building more effective decision support systems is presented in [4] and [5].

A brief overview of a six-valued logic, which is a generalized Kleene's logic, has been first presented in [19]. In [13] this logic is further developed by assigning probability estimates to formulas instead of non-classical truth values. The six-valued logic distinguishes two types of unknown knowledge values - permanently or eternally unknown value and a value representing current lack of knowledge about a state [15].

Two kinds of negation, weak and strong negation are discussed in [26]. Weak negation or negation-as-failure refers to cases when it cannot be proved that a sentence is true. Strong negation or constructable

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 242–247.

falsity is used when the falsity of a sentence is directly established.

The semantic characterization of a four-valued logic for expressing practical deductive processes is presented in [7]. In [16] it is shown that additional reasoning power can be obtained without sacrificing performance, by building a prototype software model-checker using Belnap's logic.

The four truth values in Belnap's logic are elements of an approximation lattice [7] in Fig. 1. The information about the truth-value of a sentence can have values from None to Both.

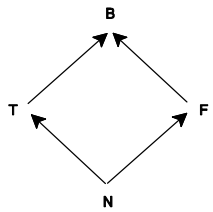


Fig.1 Approximation lattice

The four truth values are arranged in a logical lattice [7] in Fig. 2. A logical conjunction and logical disjunction are related to the meet operation and to the join operation respectively.

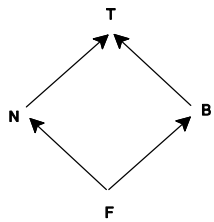


Fig.2 Logical lattice

Bi-dimensional systems representing and reasoning with temporal and uncertainty information have appeared also in [12] and [20].

Ten-valued logic was used in [23] and [24] to order default theories and distinguish different sorts of information. Ten-valued logic composed of four basic and six composed values was applied in for performing implication, justification, and propagation in combinatorial circuits.

Sixteen generalized truth values obtained as a power set of the initial truth values in Belnap's logic are arranged in a trilattice [25].

Preliminaries

Let P be a non-empty ordered set. If $sup\{x,y\}$ and $inf\{x,y\}$ exist for all $x,y \in P$, then P is called a *lattice*, [2]. In a lattice illustrating partial ordering of knowledge values, the logical conjunction is identified with the meet operation and the logical disjunction with the join operation.

A *context* is a triple *(G, M, I)* where G and M are sets and $I < G \times M$. The elements of G and M are called *objects* and *attributes* respectively [10]. The set of all concepts of the context *(G, M, I)* is a complete lattice and it is known as the *concept lattice* of the context *(G, M, I)*.

Lukasiewicz's three-valued valued logic has a third value, ½, attached to propositions referring to future contingencies. The third truth value can be construed as 'intermediate' or 'neutral' or 'indeterminate'. Lukasiewicz's generalized logic is done by inserting evenly spaced division points in the interval between 0 and 1.

Association Rules

A context *(G, M, I)* satisfies the association rule $Q \rightarrow R$ *{minsup, minconf}*, with $Q, R \in M$, if

$$sup(Q \rightarrow R) = |(Q \cup R)'|/|G| \geq minsup,$$
$$conf(Q \rightarrow R) = |(Q \cup R)'|/|Q|' \geq minconf$$

provided $minsup \in [0, 1]$ and $minconf \in [0, 1]$.

The ratios $|(Q \cup R)'|/|G|$ and $|(Q \cup R)'|/|Q|'$ are called, respectively, the *support* and the *confidence* of the rule $Q \rightarrow R$. In other words the rule $Q \rightarrow R$ has support σ% in the transaction set T if σ% of the transactions in T contains $Q \cup R$. The rule has confidence ψ% if ψ% of the transactions in T that contain Q also contain R.

Support is used for filtering out infrequent rules, while confidence measures the implication relationships from a set of items to one another.

The confidence of an association rule is a percentage value that shows how frequently the rule head occurs among all the groups containing the rule body. The confidence value indicates how reliable this rule is. The higher the value, the more often this set of items is associated together.

III. MAIN RESULTS

The generalized constructive truth-value space has as a base a set I = { T, F, t, f } containing the initial truth values

T - a sentence is constructively proven
F - a sentence is constructively refuted
t - a sentence is acceptable
f - a sentence is rejectable

The power set of I gives sixteen generalized truth values. The empty multivalue is denoted by N and A represents the set that contains the initial truth values T, F , t, f.

P(I)= [{ },{ T }, { F },{ t}, { f}, { T, F }, {T, t},
{T, f}, {F, t}, {F, f}, {t, f}, {T, F, t},
{T, F, f}, { T, t , f}, { F, t , f}, { T, F , t, f}]

The truth table for the 16 truth values is given in [25]. The three partial orderings arrange elements according to the possessed degree of information $\leq i$, truth $\leq t$ and constructivity $\leq c$ respectively. The bounds relative to the three partial orderings are shown in Table 1. Accomplished constructions (proofs and disproofs) are presented by constructive values. Non-constructive truth values do not imply any completed construction.

The lattice on Fig. 3 has five information levels, five logical levels and five levels of constructivity shown in Table 2. The sixteen truth values from the five information levels are arranged in a meet-distributive lattice Fig. 4. In such a lattice the meet of two elements of illustrates the relations between these elements that can be seen in the trilattice Fig. 4 with the advantage that if there is a meaning like $F \wedge f$ is N for two elements than this meaning is valid for all the elements at that level in Fig. 3.

Relative ordering	Bounds	Elements in P(I) being most and least
$\leq i$	A, N	informative
$\leq t$	Tt, Ff	true
$\leq c$	TF, tf	constructive

Table 1. The bounds relative to the three partial orderings

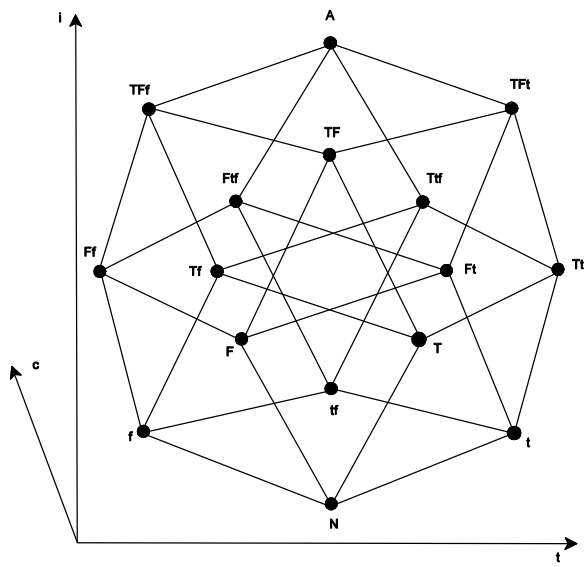


Fig. 3 Trilattice (projection i-t)

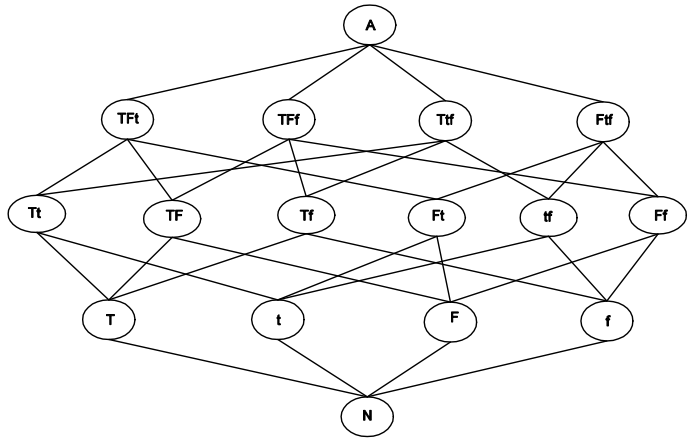


Fig.4 information meet-distributive lattice

	Information	Truth	Constructivity
1	N	Ff	tf
2	T, F , t, f	F, f, TFf, Ftf	t, f, Ttf, Ftf
3	TF, Tt, Tf, Ft, Ff, tf	A, TF, Tf, Ft, ft, N	A, Tt, Tf, Ft, Ff, N
4	TFt, TFf, Ttf, Ftf	T, t, TFt, Ttf	T, F, TFt, TFf
5	A	Tt	TF

Table 2 levels

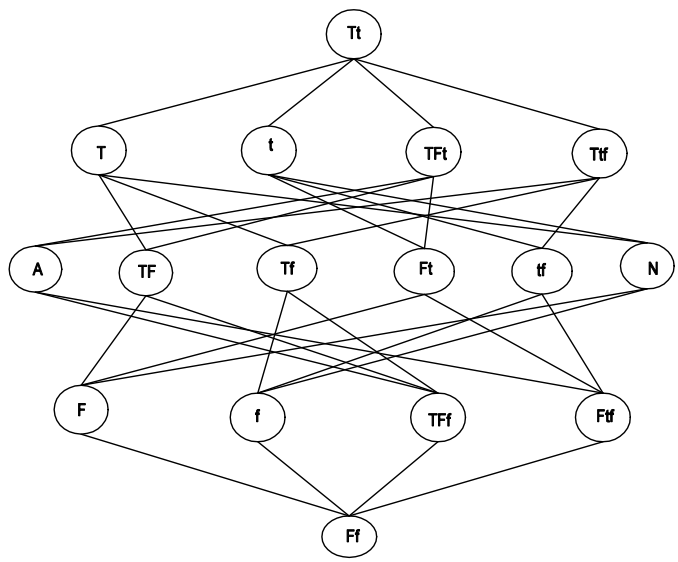


Fig. 5 Truth meet-distributive lattice

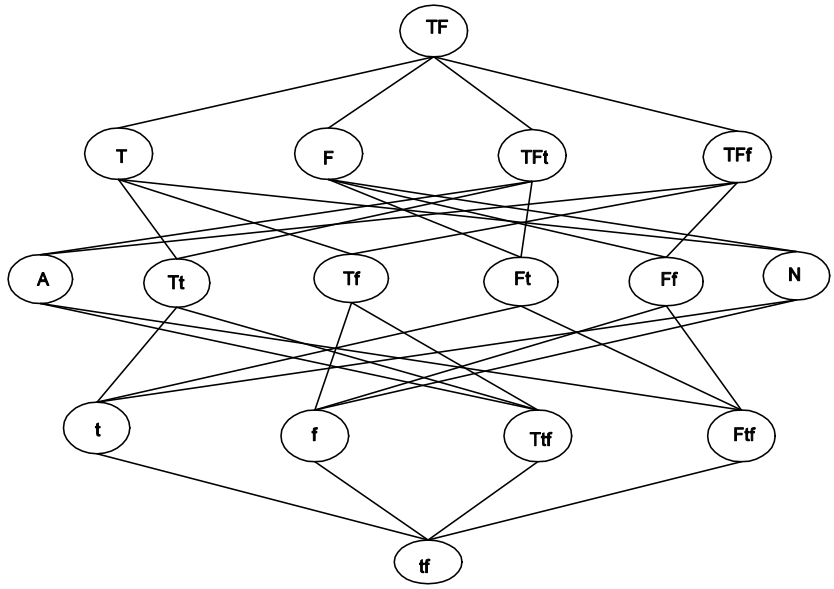


Fig. 6 Constructivity meet-distributive lattice

IV. THE SYSTEM

A number of decision support systems and even intelligent assessment systems lack the ability to reason with inconsistent information. Such a situation occurs when, f. ex. information is coming from different sources. Reasoning by applying classical logic cannot solve the problem because the presence of contradiction leads to trivialization, i. e. anything follows from 'correct and incorrect' and thus all inconsistencies are treated as equally bad.

We propose use of an intelligent assessment sub-system for comparing costumers' requirements and available offers and suggesting appropriate solutions. The intelligent agents provide expert advises to customers applying association rules as described in Section II.

Furthermore we propose application of Lukasiewicz's generalized logic for working with initial values assigned to each assessment criteria. This way the system will be able to make decisions based on different reviews and time constrains. As a result the system will be able to better facilitate the process of providing expert advises.

V. CONCLUSION

In this paper we propose use of sixteen-valued logic for dealing with incomplete and inconsistent information in intelligent tutoring systems. The five information levels, five logical levels and five levels of constructivity in trilattice of sixteen truth values are arranged in three meet-distributive lattices.

REFERENCES

[1] Agrawal, R., Imielinski, T., and Swami, A.: Mining association rules between sets of items in large databases. Proceedings of ACM SIGMOD international conference on management of data, Washington, DC, USA, (1993), 207-216

[2] Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., and Verkamo, A.I.: Fast discovery of association rules. (Uthurusamy, F., Piatetsky-Shapiro, G., Smyth, P. eds) Advances in Knowledge discovery of association rules, MIT Press, (1996), 307--328

[3] Agrawal, R. and Srikant, R.: Fast algorithm for mining association rules. Proceedings of the 20th very large data base conference, Santiago, Chile, (1994), 487-489

[4] Arnott, D. and G. Pervan, G.: A critical analysis of decision support systems research, Journal of Information Technology, 20(2), (2005), 67-87.

[5] Baskerville, R., and Myers, M.: Information Systems as a Reference Discipline, MIS Quarterly, 26(1), (2002), 1-14.

[6] Bastide, T., Taouil, R., Pasquier, N., Stumme, G., and Lakhal, L.: Mining frequent patterns with counting inference. SIGKDD explorations, Special issue on scalable algorithms, 2(2), (2000), 71-80

[7] N. .J. Belnap, "A useful four.valued logic", Modern uses of multiple-valued logic, J.M. Dunn and G. Epstain (eds), D. Reidel Publishing Co., Dordrecht, 1977, pp. 8-37

[8] Brin, S., Motwani R., Ullmann, J.D., and Tsur, S.: Dynamic itemset counting and implication rules for market basket data.

Proceedings of the ACM SIGKDD international conference on management of data, Tuscon, AZ, USA, (1997), 255-264

[9] Carpineto, C. and Romano, G.: Concept Data Analysis: Theory and Applications, John Wiley and Sons, Ltd., (2004)

[10] B. A. Davey, and H. A Priestley, "Introduction to lattices and order," Cambridge University Press, Cambridge, 2005

[11] Delgado, M., Sanchez, D., Martin-Bautista, M.J., and Vila, M.A.: Mining association rules with improved semantics in medical databases. Artificial Intelligence in Medicine, 21(1-3), (2001)

[12] Felix, P., Fraga, S., Marin, R., and Barro, S.: Linguistic representation of fuzzy temporal profiles, International Journal of Uncertainty, Fuzziness and Know ledge Based Systems, 7(3), (1999), 243-257

[13] Fitting, M.: Kleene's Logic, Generalized, Journal of Logic and Computation, 1(6), (1991), 797-810

[14] Ganter, B., Stumme, G., and Wille, R.: Formal Concept Analysis - Foundations and Applications. Lecture Notes in Computer Science, Springer Verlag, 3626, (2005)

[15] Garcia, O. N., and Moussavi, M.: A Six-Valued Logic for Representing Incomplete Knowledge, Proc. of the 20th International Symposium on Multiple-Valued Logic (ISMVL), IEEE Computer Society Press, Charlotte, NC, USA, May (1990), 110-114.

[16] Gurfinkel, A. and Chechik, M.: Yasm: Model-Checking Software with Belnap Logic. Technical Report 470, University of Toronto, April, (2005)

[17] Malerba, D., Lisi, F.A., Appice, A. and Sblendorio, F.: Mining spatial association rules in census data: a relational approach. Proceedings of the ECML/PKDD'02 workshop on mining official data, University Printing House, Helsinki, (2002), 80-93

[18] Merceron, A. and Yacef, K.: A Web-based Tutoring Tool with Mining Facilities to Improve Learning and Teaching. Proceedings of 11th International Conference on Artificial Intelligence in Education. F. Verdejo and U. Hoppe (eds) Sydney, IOS Press, (2003)

[19] Moussavi, M., and Garcia, N.: A Six-Valued Logic and its application to artificial intelligence, Proc. of the Fift Southeastern Logic Symposium, IEEE Computer Society Press, UNC-Charlotte, NC, USA, (1989).

[20] Mulsliner, D. J., Durfee, E. H., and Shin, K. G.: CIRCA: A cooperative intelligent real-time control architecture. Trans. on Systems, Man and Cybernetics, 23(6), (1993), 1561-1574

[21] Pasquier, N., Bastide, T., Taouil, R., and Lakhal, L.: Efficient mining of association rules using closed itemset lattices. Journal of Information Systems, 24(1), (1999), 25-46

[22] Pecheanu, E., Segal, C. and Stefanescu, D.:Content modeling in Intelligent Instructional Environment. Lecture Notes in Artificial Intelligence, Springer-Verlag, Berlin Heidelberg New Jork, 3190, (2003), 1229-1234

[23] Sakama, C.: Ordering default theories. Proceedings of the 18th International Joint Conference on Artificial Intelligence, Morgan Kaufmann, Los Altos, CA, (2003), 839-844

[24] Sakama, C.: Ordering default theories and nonmonotonic logic programs Theoretical Computer Science, 338(1-3), (2005), 127-152

[25] Shramko, Y., Dunn, J.M. and T. Takenaka, T., The Trilattice of Constructive Truth Values. Journal of Logic Computation 11(6), (2001) 761-788

[26] Wagner, G.: Vivid Logic: Knowledge Based reasoning with two kinds of negation, Lecture Notes in Artificial Intelligence, 764, Springer-Verlag, Berlin Heidelberg New York, (1994)

[27] R. Wille, "Concept lattices and conceptual knowledge systems," Computers Math. Applic., 23(6-9), 1992, 493-515

[28] Zaki, M.J.: Generating non-redundant association rules. Proceedings of the 6th ACM SIGKDD international conference on knowledge discovery and data mining, Boston, USA (2000), 34-43

[29] Zaki, M. J. and Hsiao, C. -J.: CHARM: An efficient algorithm for closed itemset mining. Proceedings of the 2nd SIAM international conference on data mining, Arlington, VA, USA, (2002), 34-43

Analog Computer to Solve Third-Order Linear Differential Equation

T. ElAli, S. Jones, F. Arammash, J. Biotidara, T. Oluwafemi
Department of Physics and Engineering
Benedict College
1600 Harden Street
Columbia, SC 29204

Abstract

An analog computer was designed and tested to solve any third order constant-coefficients and linear differential equation. The analog computer was built using operational amplifiers, resistors and capacitors. Using the Multisim simulator, various input types were tested across the input terminals of the analog computer and the results were recorded.

I. INTRODUCTION

Our goal is to build a generic electronic circuit to solve a generic 3rd order differential equation with any input.
Consider the generic differential equation to be solved

(1)

x(t) is the forcing function (the input to the system represented by this differential equation) and y(t) is the solution (the output of the same system). The variables a, b, c, and d are some real constant numbers. [1]
In the last equation, (assuming zero initial conditions) let

$$\begin{cases} y_1(t) = y(t) \\ y_2(t) = \dfrac{d}{dt} y(t) \\ y_3(t) = \dfrac{d^2}{dt^2} y(t) \end{cases}$$

Thus we have the set of three first order differential equations

$$\begin{cases} \dfrac{d}{dt} y_1(t) = y_2(t) \\ \dfrac{d}{dt} y_2(t) = y_3(t) \\ \dfrac{d}{dt} y_3(t) = -\dfrac{d}{a} y_1(t) - \dfrac{c}{a} y_2(t) - \dfrac{b}{a} y_3(t) + \dfrac{1}{a} x(t) \end{cases} \qquad (2)$$

II. METHODS

Consider the Operational amplifier circuit shown in Figure 1. The input-output relationship is given as

$$y(t) = -A \frac{1}{RC} \int x_1(t)dt - B \frac{1}{RC} \int x_2(t)dt \qquad (3)$$

In Figure 1, the output y(t) is the integral of the input arriving at the negative terminal of the Operational Amplifier. Thus the negative of the derivative of y(t) is located at the negative terminal of the Operational Amplifier. [2,3,4]

If we set RC=1 in equation (3) we will have

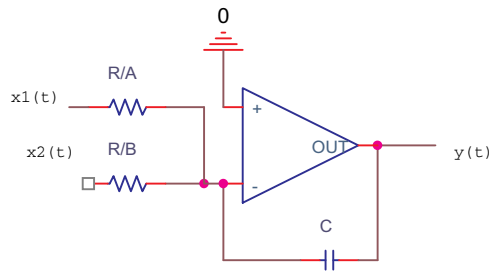


Fig. 1. Operational Amplifier Circuit

$$y(t) = -A \int x_1(t)dt - B \int x_2(t)dt \qquad (4)$$

One final step before we attempt to implement Equation (4), the solution of a generic 1st order linear constant coefficient differential equation. Consider the circuit given in Figure 2. The input-output relationship is

$$y(t) = -\frac{R_f}{R} x(t) \qquad (5)$$

You also can see that if R_f = R then we have pure inversion (unity gain). The circuit containing an inverter and an integrator connected in series can solve the differential equation given in (6). Figure 3 is a typical example of such a circuit.

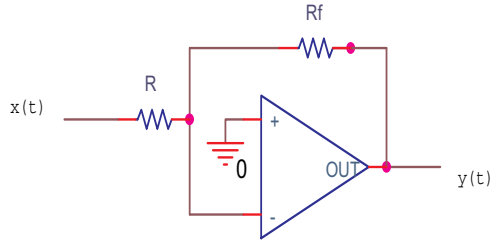


Fig. 2. Inverter

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 248–250.

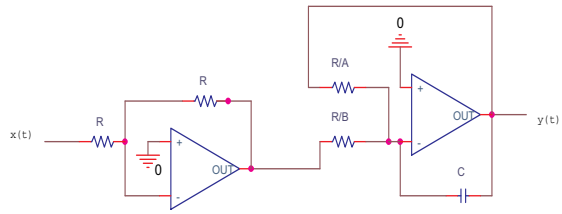


Fig. 3. A Circuit to Solve Equation 6

The circuit in Figure 3 would solve any first order differential equation of the form

$$y'(t) + Ay(t) = Bx(t) \qquad (6)$$

Knowing how to solve equation (6) is helpful in solving the set of the three coupled equations in (2) with y(t) being the output in the third-order differential equation given in (1).

In building a circuit to solve the given differential equation in (1) we will use the set of equations in (2). We have tried step input, impulse input, and sinusoidal input. All worked nicely. Next we present the step response for different real coefficients and consider critically damped, over damped and oscillatory cases.

For the over damped case we selected a=1, b=6, c=11, and d=6. The differential equation is then

$$\frac{d^3}{dt^3}y(t) + 6\frac{dt^2}{dt^2}y(t) + 11\frac{d}{dt}y(t) + 6y(t) = u(t)$$

where u(t) is the step unit signal. The eigenvalues are then at

-3.0000
-2.0000
-1.0000

Using Matlab®, the solution is plotted as seen in Figure 4.

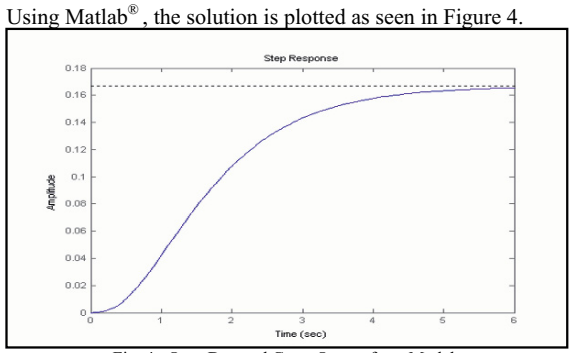


Fig. 4. Over Damped Case: Output from Matlab

The circuit to solve the third-order differential equation with the given constant values and with the input u(t) is shown in Figure 5. The simulated output is shown in Figure 6.

For the critically damped case we selected a=1, b=3, c=3, and d=1. The differential equation is then

$$\frac{d^3}{dt^3}y(t) + 3\frac{dt^2}{dt^2}y(t) + 3\frac{d}{dt}y(t) + y(t) = u(t)$$

The eigenvalues are then at

-1.0000
-1.0000
-1.0000

Using Matlab, the solution is plotted as seen in Figure 7. Using the circuit, the simulation is shown in Figure 8.

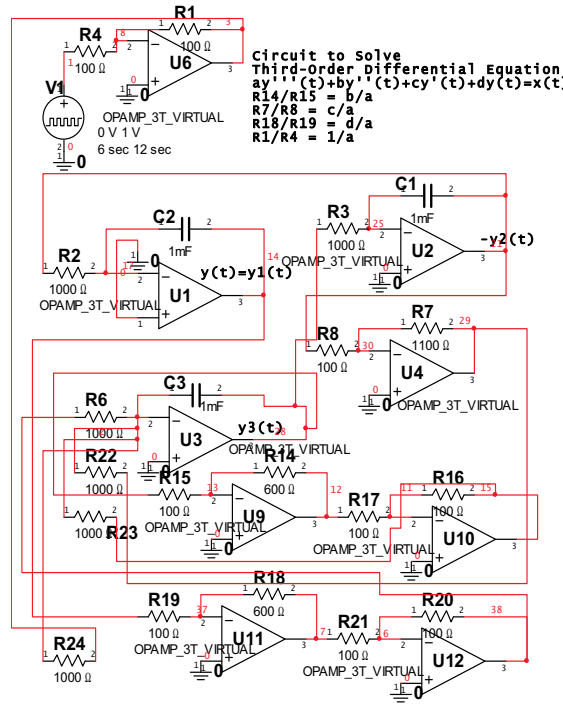


Fig. 5. Circuit to Solve the Differential Equation

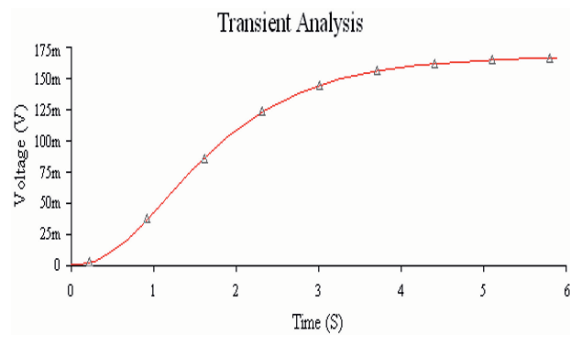


Fig. 6. Over Damped Case: from Circuit Simulation

For the oscillatory case we selected a=1, b=1, c=1, and d=1. The differential equation is then

$$\frac{d^3}{dt^3}y(t) + \frac{dt^2}{dt^2}y(t) + \frac{d}{dt}y(t) + y(t) = u(t)$$

The eigenvalues are then at
-1.0000
-0.0000 + 1.0000i
-0.0000 - 1.0000i
Using Matlab, the solution is plotted as seen in Figure 9. Using the circuit, the simulation is shown in Figure 10.

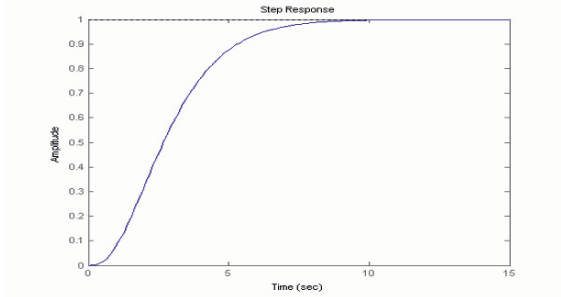


Fig. 7. Critically Damped Case: Output from Matlab

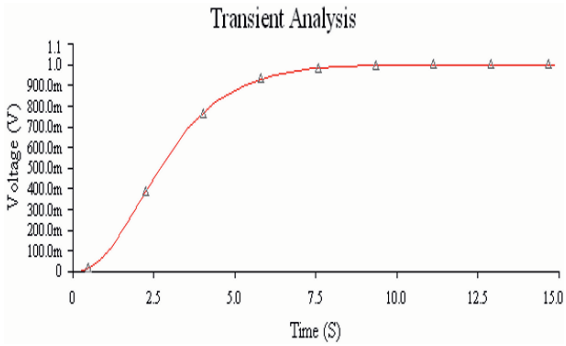


Fig. 8. Critically Damped Case: from Circuit Simulation

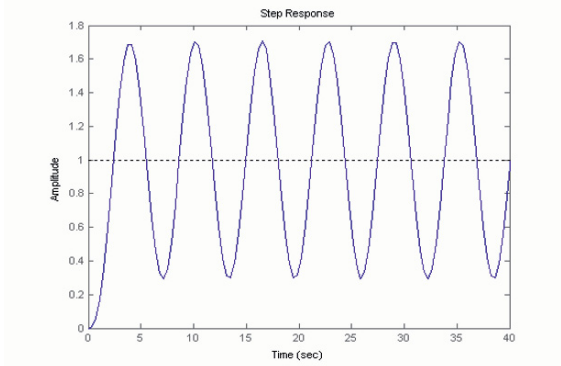


Fig. 9. Oscillatory Case: Output from Matlab

In the process of simulating the circuit we had to adjust the resistor values to accommodate the change in the coefficients. The input x(t) has to be adjusted too to accommodate the necessary time required to show enough time for the transients to settle.

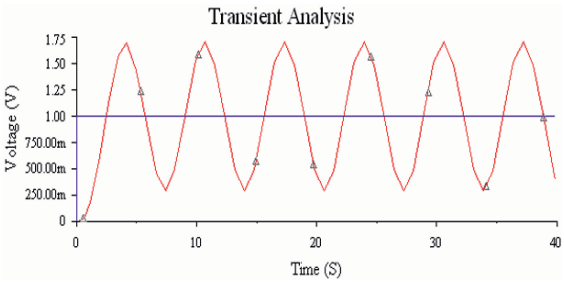


Fig. 10. Oscillatory Case: from Circuit Simulation

IV CONCLUSION

It would be discovered by looking at the graphs and also by comparing these results with what was derived analytically that the circuits worked as desired. The differential equation was solved and its outputs were a solution to the given input. In the future we will attempt to solve higher order differential equations. Practically, to solve any fourth order differential equation with any arbitrary coefficients requires a huge set of resistive values. However, since the constant values of a, b, c, and d can be translated to ratios of resistor values that makes things easier. Issues related to amplifier saturation should also be studied. [5]

V REFERENCES

[1] Henry Edward, "Elementary Differential Equations", 4th edition, *Prentice Hall*, 2000.
[2] Robert Boylestad, "Electronic Devices and Circuit Theory", 8th edition, *Prentice Hall*, 2002.
[3] J. W. Nilson, "Electric Circuits", 7th edition, *Prentice Hall*, 2005.
[4] J. W. Nilson, "Introduction to Pspice Manual using Orcad", 7th edition, *Prentice Hall*, 2005.
[5] C. Chen, "Analog & Digital Control System Design", 1st edition, *Saunders*, 1993.

Data Processing for Mapping in Mobile Robotics

Tomas Neuzil, Ondrej Jez

Brno University of Technology
Faculty of Electrical Engineering and Communication
Department of Control and Instrumentation
Kolejni 4, 62100, Brno
Czech Republic
ondrej.jez@phd.feec.vutbr.cz
neuzil@feec.vutbr.cz

Abstract: **Mobile platforms are usually equipped with sensors for obstacle avoiding or building maps of area. In this article there is presented an algorithm based on Hough transformation. Also experiment for determining the properties of used sensor is presented.**

1. Introduction

This article deals with two related topics. It describes two parts of localisation algorithm. The first part deals with the local map processing based on Hough transformation. In the second part, the parameters of a SICK laserscanner are tested and described. These results serve as background in sensor modeling and at the same time as sensor verification for 6 DOF SLAM.

In the present day, there is high accent on research in the area of the autonomous mobile robotics. Autonomy in the robotics is taken as a capability to pursue activity without human intervention in an unstructured area. The mobile robot has to fulfill following conditions in order to be autonomous.

- The ability to work for long time without human intervention.
- The ability to maneuver through working area.
- The ability to avoid situations that could lead to destruction of the robot or the injury.
- The ability to collect information about its working area.

For all of the above mentioned conditions is necessary to know the precise position of the vehicle. [5]

It is possible to obtain the information about a robot's position in several ways. The most interesting method of measurement is measurement without intervention to the measured area which assumes impossibility to make artificial landmarks. The way how to fulfill these strict conditions is the use of rangefinder, either optical, ultrasonic or microwave [1]. Accessibility of industrial laser scanners allows their expansion in mobile robotics for the purposes of navigation and position determination.

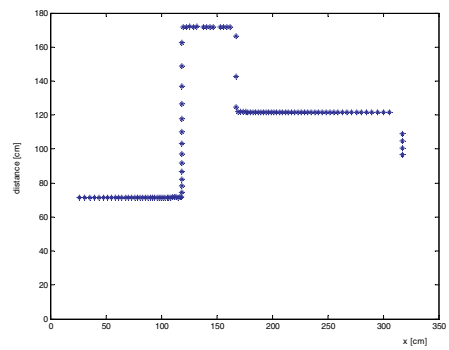


Figure 1 – Depth image of the observed space

2. Proximity Sensor Data Processing

The scanner outputs data which are saved in a matrix form. This matrix contains coordinates of single points belonging to obstacles in the space [2], [3].

The set of data corresponding to one measurement is very large. Therefore it is necessary to transform the space described by coordinates of obstacle points to an efficient form. One of the ways is to describe the obstacles by parts with curves. This "description" transforms the data stored in a sensoric map (raw sensor data) to a geometric map (data presented as set of geometric primitives).

Hough Transform is one of the methods for finding curves hidden in larger amounts of data (fig. 1). It is mostly used in image processing. In our case we use the Hough transform for finding straight lines in the depth image. The main advantage of the Hough transform is that it can detect lines or curves, even if they are broken. It means the measured data does not have to be complete to be processed by Houh transform. Each point in Hough accumulator (fig. 2) corresponds to a line at angle φ and distance d from the origin in the data space. The value of a function in Hough space gives the point density along the line in the data space.

We consider all the lines which go through that point at a set of angles for each point in the original space. For each angle φ,

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 251–254.

calculate the distance d to the line through the point at that angle.

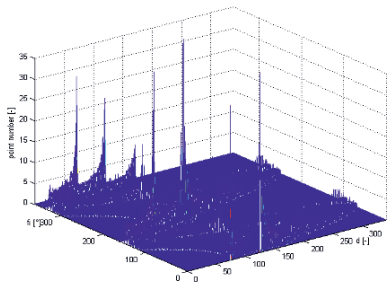


Figure 2– Hough Accumulator

For each line we increment a count (initialized at zero) in the matrix called accumulator at coordinate (d, φ).

The number in the cell of the accumulator corresponds to the number of points in the original space that belongs to the straight line with coefficients (d, φ). After processing all the lines through all the points of the original space, the cells of the accumulator with the highest values correspond to the lines in the depth picture.

The main problem is that points in the depth picture can represent more than one line. This fact conducts to the need of mechanism how to find all lines in the picture. We can use a simple method of threshold, where the value of a threshold represents the number of points belonging to the corresponding line. The main disadvantage of this method is that all lines above the threshold are found (we are looking for local maximums). If the depth picture contains short lines (with small number of points), poor setting of the threshold value can ignore these short lines, even if we want them to be processed.

In our algorithm we use basics of cluster analysis for finding the maximums of the Hough accumulator [3], [4]. The first step is removing of lines with small number of points on it. This is done by easy threshold setting. The results are clusters of points. Each cluster represents one line.

Figure 3 depicts filtered Hough accumulator. In one cluster there are all lines with similar parameters included.

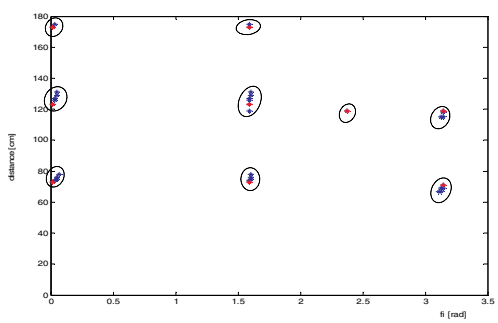


Figure 3– Objects separated into clusters

The basic principle of clustering method is to find the dissimilarity between every pair of objects (points) in the data set. Each object is described by the *m* characteristic parameters:

$$o_h = (x_{h1} \ldots\ldots x_{hm}) \tag{1}$$

each object has two parameters:

$$x_{h1} = d_h$$
$$x_{h2} = \varphi_h \tag{2}$$

The dissimilarity function has following properties:

$$d(o_h, o_s) = 0 \Leftrightarrow o_h = o_s$$
$$d(o_h, o_s) \geq 0 \tag{3}$$
$$d(o_h, o_s) = d(o_s, o_h)$$

The most frequently used method for determining the dissimilarity is Euclidean metric:

$$d(o_h, o_s) = d_E(o_h, o_s) = \sqrt{\sum_{j=1}^{m} (x_{hj} - x_{sj})^2} \tag{4}$$

We can classify objects into groups – clusters in the moment when we have counted all dissimilarities between all objects,. A cluster is defined as a continuous group of objects. An object o_h is continuous with o_l object for a given threshold T if there exists series of objects:

$$o_h = o_1, o_2, o_3 \ldots, o_k = o_l \tag{5}$$

$$d(o_i, o_{i+1}) \leq T \text{ for i=1,2…. k-15)} \tag{6}$$

In the moment we have all accumulators classified into clusters, we can find object with the maximum value. These found objects represent the straight lines $l_1 \ldots l_N$ with parameters (d_1, φ_1..... d_N, φ_N). These lines are then used for approximation of measured points.

The Hough transform allows finding straight lines in the depth picture. The boundaries of single line segments have to be found in the next step of the algorithm. The possible ways how to find these boundaries is counting distances of individual points from the line and then determining if the point is a member of the abscissa or not.

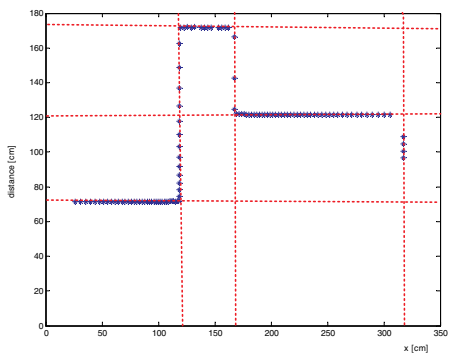


Figure 4 – Depth image of the observed space with found lines

3. Sensor description

Processed data e.g. lines equations, lines intersections or point of symmetry are used as an input for the localization algorithm. This algorithm is based on the Kalman filter.

$$x_{(k+1)} = Fx_{(k)} + Gu_{(k)} + v_{(k)}$$
$$y(k) = H_{(k)} + w_{(k)} \quad (7)$$

These equations describe mobile robot with sensor as dynamic system. In both equations there are noise vectors $v_{(k)}$ an $w_{(k)}$.

These vectors are important values influencing accuracy of the position estimation. The value of $v_{(k)}$ depends on the quality of the mathematical model of the robot. Kinematic model utilizes vectors of velocity for computing position while dynamic model deals with all forces that take effect in the robot motion.

Experiments were focused on the measurement of noise $w_{(k)}$. The mobile robotic platform is equipped with the SICK LMS 200 laser scanner.

In order to validate the influence of an obstacle's reflectivity on the measurement reliability, it was necessary to manipulate the reflectivity of certain material. Since a diffuse material was used and the laser scanner's wavelegth is constant, the reflectivity could only be changed by the change of material properties. This was done by using a matte paper and printing different shades of gray, controlling the percentage of the black in coverage . To limit the number of measurements to be taken at different distances, the number of shades was set to 5, where 5 corresponds to white paper and thus highest reflectivity and 1 corresponds to 100% coverage by black ink. Although the measurement of the reflectivity was not available in the time of the experiment, the reflectivity of the shade No. 3 is set to 18%. This value is well known in digital photography as a 18% middle grey. It lies in the middle of the black-white gamut. Its composition in [R,G,B] space is defined as [127,127,127]

Another issue was to estimate whether the reliability of the measurements changes over the distance differently for objects of different colors. The fact that objects seem to have different colors in human perception is due to the sensitivity of the object's reflectivity to the incidental radiation wavelength, meaning that certain objects have higher reflectivity for certain wavelengths (colors). To perform such experiment, a set of colored papers was created. Unfortunately this part of the experiment lacks important validation of the sample objects because a device to measure spectral intensities was not available at the time of experiment.

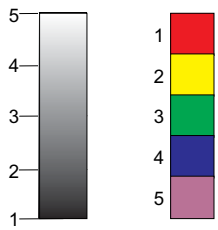


Figure 5 Tested colors and grey levels

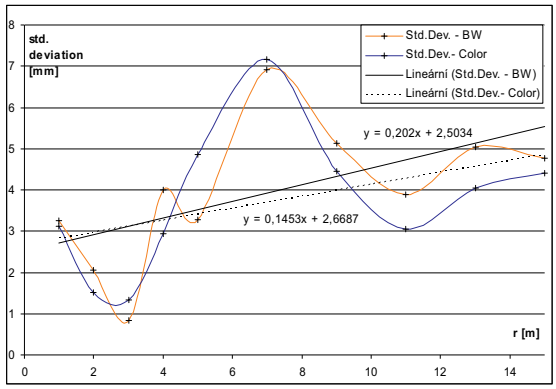


Figure 6 Standard deviation vs. distance

Therefore the results are only illustrative and they cannot be used for modeling of the sensor taking color sensitivity into account.

In figure 6 the dependency of the standard deviation on distance measurement is presented. Values of std. deviation are average values through all grey levels resp. all colors.

Based on these results, model of the sensor was created.

$$\begin{bmatrix} r_i \\ \gamma_i \end{bmatrix} = \begin{bmatrix} \sqrt{x_i^2 + y_i^2} \\ tg\left(\frac{y_i}{x_i}\right) \end{bmatrix} + \begin{bmatrix} w_r \\ w_\varphi \end{bmatrix} \quad (8)$$

This simple model is taken with respect of the sensor coordinates. In case of use this model for mapping or lacalisation it is necessary to transfer it into global coordinate frame $[X_G,Y_G]$.

From our experiment we are able to determine the componet of measurement noise w as:

$$w_r = 1.12x + 9.24 \quad (9)$$

The other component w_φ was determined on the basis of the product datasheet [7] as: $w_\varphi = 1°$.

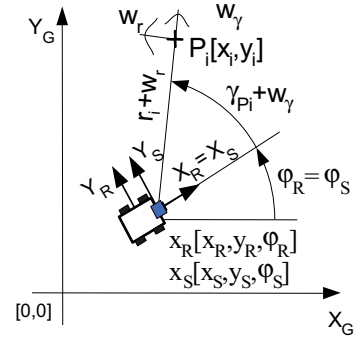


Figure7 Model of the sensor

4. Conclusion

In this time the parts with robot model and real mobile platform are accomplished. Software with implemented Hough transformation and cluster analysis is also ready and tested on the sensor model data. The future work is to combine singular parts together and make compact system for Simultaneous Localization and Mapping in 2D areas. Also practical tests of algorithms are planned. In this time all the algorithms are programmed in Matlab, so they are not determined for realtime use.

5. Acknowledgements

This project was supported by grant MSM0021630529 – Intelligent systems in automation.
This work was supported by the Ministry of Education of the Czech Republic under Project 1M0567.

6. References

[1] Blais,F. : A review of 20 Years of Range Sensor Development, Proceedings of SPIE-IS&T Electronic Imaging, SPIE VOL. 50123, pp. 62-76
[2] Herbert, M.: Active and Passive Range Sensing for Robotics, Carneige Mellon University, The Robotics Institute, Research Report
[3] Tina Yu Tian, Shah M.: Recovering 3D Motion of Multiple Objects Using Adaptive Hough Transform, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 10, October 19
[4] Žák, L.: Shluková Analýza, Automatizace, Vol. 3, March 2004
[5] Michael Csorba : Simultaneous localization and map building, Disertation Thesis, University of Oxford, 1997
[6] www.sick.com (1.9.2007)

Piecewise Continuous Systems Used in Trajectory Tracking of a Vision Based X-Y Robot

Haoping Wang, Christian Vasseur and Vladan Koncar*
Laboratoire d'Automatique, Génie Informatique & Signal (LAGIS - UMR CNRS 8146)
Université des Sciences et Technologies de Lille, France

*GEnie & Matériaux TEXtiles, ENSAIT, France
haoping.wang@ed.univ-lille1.fr - christian.vasseur@univ-lille1.fr - vladan.koncar@ensait.fr

***Abstract*- This paper deals with a new approach of trajectory tracking for a vision based x-y robot which delivers a delayed and sampled output. A control theory of using a class of piecewise continuous systems named as piecewise continuous controllers and a specific observer enabling sampled tracking is developed. The experimental results show the effectiveness and robustness aspects of the method.**

I. INTRODUCTION

Nowadays, Computer Numerical Controlled (CNC) tools are widely used in manufacturing automation, and the two-axis motion CNC system is the mainstream configuration [7]. Compare to [9, 13] using the neural or fuzzy based control, one presents here a class of hybrid systems called Piecewise Continuous Systems (PCS) to realize a vision based x-y robot tracking control.

This kind of systems which is firstly introduced in [6] and then developed in [2, 12] is characterized by its two input spaces and two time spaces. Each input corresponds to one specific time space. The first refers to the discrete time space $S = \{t_k, k = 0, 1, 2, \cdots\}$ called switching space. The second refers to the continuous time space $t \in \Im - S$ with $\Im = \{t \in [0, \infty]\}$. At each switching instant, the plant is controlled from the first input space V^σ. And between two switching instants, the plant is controlled from the second input space U^r.

And two successive switching instants t_k and t_{k+1} delimit a piece, noted $\Phi_k : \left\{\Phi_k \,\middle|\, \forall t \in]t_k, t_{k+1}[\right\}$.

Thus a state-space based approach uses four vector spaces: *n*-dimensional state space denoted Σ^n, *m*-dimensional output space denoted Y^m and the preferred two input spaces (V^σ & U^r) is well adapted to represent PCS dynamics in the Φ_k piece as follows:

$$x(t_k^+) = z(t_k, v(t_k)), \quad \forall t_k \in S \tag{1a}$$

$$x(t) = f(t_k, t, x(t_k^+), u_{]t_k, t[}), \quad \forall t \in \Im - S \tag{1b}$$

$$y(t) = h(t, x(t)), \quad \forall t \in \Im \tag{1c}$$

In reference to the classification of Tittus and Egardt [10], this class of control systems has hybrid properties and extends the concept of compound control proposed by Laurent [8] and Vasseur [11]. Moreover, according to Branicky's taxonomy of hybrid systems [1], these control units are characterized by autonomous switching and controlled impulses.

These systems, characterized by autonomous switchings and controlled impulses can be used as regulators: Piecewise Continuous Controllers (PCC). Though the standard PCC [6] requires a linear model of the plant to be controlled, it is shown here that an adaptation of the PCC gives rise to a particular regulator: called Derived PCC (DPCC) that allows control without knowledge of the plant's model. The DPCC are easily implemented on digital calculators and allow trajectories tracking by the plant's output.

And during the last few years, there has been a considerable amount of interest in the control of vision based mechanical systems. The interest comes from the need of supervision in remote control especially via Internet based network, more flexible contactless wiring and improved signal/noise ratio. But the problem is that the introduction of these sensors, often deliver sampled and delayed signals due to their digital nature and computation-transfer time (image processing) respectively. The challenge here is to consider a low cost CCD camera as a contactless sensor to realize a robust trajectory tracking.

Analyzing the difficulties of previous vision based research works [4-5]; it seemed that the camera signal has not been sufficiently exploited. Therefore, the efforts have been focused on the development of an accurate observer using the PCS theory. This developed Piecewise Continuous reduced order Luenberger Observer (PCO) which consists mainly on two PCS and a Reduced Order Luenberger Observer (ROLO) gives a precise x-y robot's position estimation.

This paper is organized as follows: Section II presents the modeling of a vision based x-y robot system. Then in section III from a PCC for a linear system, one develops a particular controller: DPCC. After that to estimate the vision based robot's position, one develops a particular observer: PCO in section IV. And the x-y robot's tracking experimental results are presented in section V. Finally, one concludes with some remarks in the section VI.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 255–260.

II. Modeling X-Y Robot

A. Experimental platform

The vision based x-y robot experimental platform which is presented in Fig. 1 contains the next four parts:

1) The mechanical system is composed of an X-Y aluminum chassis, enabling only 48 cm displacement on each. The x-axis moves along the y-axis (cross LM guided). Axes are actuated by AC servo motors via notched belt.

2) The controller is implemented on the dSpace based Digital Signal Processing card DS1103 via ControlDesk integrated with Matlab/Simulink. The control signals are sent to a power amplifier via ± 10 V DAC.

3) Two servo motors (SANYO DENKI PY2A015A3-A3) and two AC motors (P50B050020-DXS00M) are driven by a dSpace computer input/output card via a power amplifier supplied with 240 V. The AC motors delivers a nominal couple of 3.0 Nm with a power of 200 W. The platform returns the cart's continuous position through encoders.

4) The vision system of a low cost IR CCD (Jai M50 IR) camera with a sampling rate of 25 frames/sec and a low resolution of 640 × 480 pixels in non-interlaced mode. This camera is linked to a vision computer which constitutes an image acquisition card ELTEC PC-EYE 4 and an image processing software TEKVIS.

B. Modeling the x-y robot

On x-axis, the motor-cart model considers the motor terminal voltage u_x as its input and the x-y robot horizontal displacement x_C as its output. Thus consider [3], the dynamic system can be modeled as follows:

$$\ddot{x}_C = (-\dot{x}_C + k_x u_x)/\tau_x \tag{2}$$

where: k_x the overall gain of the motor-cart on x-axis. τ_x the time constant of the motor-cart on x-axis.

In the same way, we can modelize the motor-cart on y-axis as: $\ddot{y}_C = (-\dot{y}_C + k_y u_y)/\tau_y$.

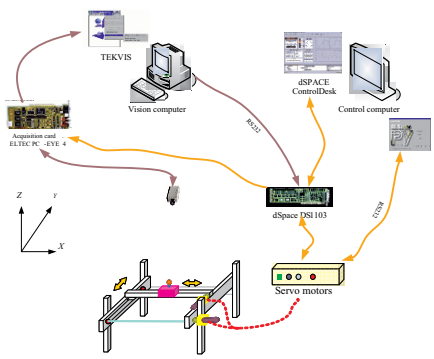


Fig. 1. Vision based x-y robot (system for disposal)

C. Modeling the vision based signals

This vision system is used to determine the x-y robot's position (x_C, y_C) coordinates and transmits them to the control computer via the RS-232 serial communication port. The measurements of the camera are available at a sampling rate of $T = 38$ ms (acquisition-processing-transfer time). And in order to facilitate the vision sensing, an infrared LED has been added on the upper tip of fixed pole on the x-y robot as illustrated in Fig. 1. In the first phase and in order to synchronize the camera with the control algorithm, the camera is triggered by an external periodical pulse signal, generated via the dSpace card with a sampling period equal the acquisition-processing-transfer time.

As soon as the control computer receives the fixed pole's coordinates, a four step "TSAI" calibration method [4] is carried out in the control system. It supports not only the real x-y robot's pole position but also the compensation of the deformations caused by the camera's lenses.

Under this condition, the sampled and delayed measurements of x-y robot are available:

$$(x_C(kt_e - T_e), y_C(kt_e - T_e)) = (x_{C,k-q}, y_{C,k-q}) \tag{3}$$

with: t_e the camera's sampling cycle, and $T_e = qt_e$, where: k, q are integers. T_e represents the delay time corresponding the time necessary for data acquisition, processing and transfer.

In the vision system, t_e=38 ms and q=*1*. Finally from (2), the sampled and delayed $(x_{C,k-1}, y_{C,k-1})$ are computed.

In the reason of the x-y robot structure's symmetry and the simplification, one presents only the x-axis control. With the stale representation, equations (2-3) can be written as:

$$x'(t) = A.x(t) + B.u(t) \tag{4}$$

$$y(t) = C.x(t) \tag{5}$$

$$\theta_{k-1} = x_{C,k-1} \tag{6}$$

with: $A = \begin{bmatrix} 0 & 1 \\ 0 & -1/\tau_x \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ k_x/\tau_x \end{bmatrix}$ and $C = [1 \quad 0]$, and $x(t) = [x_C(t) \quad \dot{x}_C(t)]^T$, $u(t) = u_x(t)$, $y(t)$ and θ_{k-1} representing respectively the state, the input, the output and the camera's delayed and sampled measurements of x-axis motor-cart.

III. Piecewise Continuous Control

A. Piecewise continuous controller

In reference to equations (1) and according to the modeling procedure, the behavior of a PCC for a linear system, denoted as $\Sigma(\{kt_e\}, \alpha, \beta, \delta, \gamma)$ can be summarized as follows:

The PCC's state $\lambda(t) \in \Sigma^{\hat{n}}$ is switched to forced values at regular intervals of period t_e, the same of the camera's sampling time. And the corresponding switching set is represented by $S = \{ kt_e, k = 0,1,2,... \}$.

The equations of the controller in the Φ_k piece are:

$$\lambda(k.t_e^+) = \delta.\psi(k.t_e)\ ,\ k \in S \quad (7a)$$

$$\lambda'(t) = \alpha.\lambda(t) + \beta\varphi(t)\ ,\ t \in \Im - S \quad (7b)$$

$$w(t) = \gamma.\lambda(t)\ ,\ t \in \Im \quad (7c)$$

- Equation (7a) defines the controller's state at switching instants, by means of a first input $\psi(k.t_e) \in V^s$ upon the discrete time space *S*, and according to the linear relationship characterized by the matrix $\delta \in \Re^{\hat{n} \times s}$.
- Equation (7b) describes the continuous evolution of the controller's state $\lambda(t) \in \Sigma^{\hat{n}}$ under the second space $\varphi(t) \in U^{\hat{r}}$ upon on the continuous time space $t \in \Im - S$. $\alpha \in \Re^{\hat{n} \times \hat{n}}$ and $\beta \in \Re^{\hat{n} \times \hat{r}}$ being the state matrix and the input control matrix of the controller. For a simplified PCC, β can be imposed zero. Thus the only parameter that defines the behaviors of the controller's state in this interval of time is α which can take an arbitrary value. Usually, it is fixed such that the PCC is stable between switching instants.
- Equation (7c) is the output equation of the controller, characterized by the matrix $\gamma \in \Re^{\hat{m} \times \hat{n}}$. The output $w(t) \in Y^{\hat{m}}$ constitutes the input command to be fed to the plant.
- Fig. 2a, Fig. 2b and Fig. 2c gives respectively a PCC's realization diagram, symbolic representation and its state's evolution. It's important to note that, normally $\lambda(k.t_e^-) \neq \lambda(k.t_e^+)$, according to (7a, 7b).

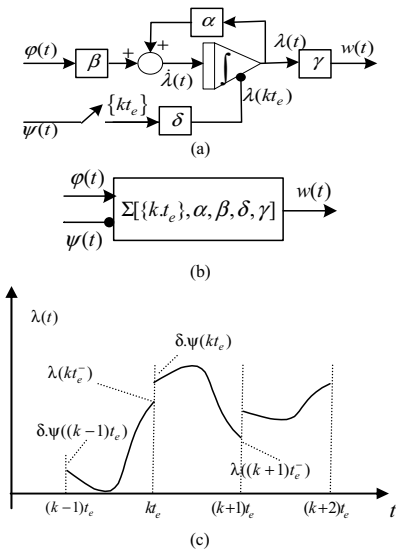


Fig. 2. Piecewise Continuous Controller

Note that from now on, the discrete values of every function will be considered as being sampled at t_e period and to simplify the notations, any time function $f(t)$ at a given $k.t_e$ instant will be written as $f(k.t_e) = f_k \quad t \in S$. Moreover, if any signal $f(t)$ is discontinuous, we shall consider the right value at the discontinuity since the switching instant at each $k.t_e$ imply consequences occurring at every $k.t_e^+$. However, for simplification sake, the notation f_k will be used, instead of the strict one: $f_k^+ = f(k.t_e^+)$.

B. Control strategy

The aim is to define the PCC parameters $\psi(t)$ and δ so as to achieve discrete tracking of a state trajectory $c_s(t)$ by the plant's state $x(t)$ at each switching instant and with one sampling period delay:

$$x((k+1)t_e) = c_s(k.t_e)\ ,\ k = 0,1,2,... \quad (8)$$

To do so, we consider the PCC and the x-axis system as a closed loop system whose equations are as follows:

$$x'(t) = A.x(t) + B.u(t)\ ,\ t \in \Im \quad (9)$$

$$\lambda_k = \delta.\psi_k\ ,\ k \in S \quad (10)$$

$$\lambda'(t) = \alpha.\lambda(t)\ ,\ t \in \Im - S \quad (11)$$

$$u(t) = \gamma.\lambda(t)\ ,\ t \in \Im \quad (12)$$

By integration, the first three equations allow us to write in a sampled format, the next step value x_{k+1} of the state as a function of its previous one x_k:

$$x_{k+1} = f.x_k + M.\lambda_k \quad (13)$$

with $f = e^{A.t_e}$ and $M = f.\int_0^{t_e} e^{-A\tau} B.\gamma.e^{\alpha\tau} d\tau$.

In order to realize the discrete tracking which is defined in (8), we only have to fix down the tracking condition $x_{k+1} = c_{s,k}$ where $c_s(t)$ is the desired state trajectory. Thus, from (13) we have

$$\lambda_k = M^{-1}\left\{c_{s,k} - f.x_k\right\} \quad (14)$$

Equation (14) gives the switching value of the controller's state, under the condition that M^{-1} exists [6]. Hence, in this case, we are able to define the PCC with:

$$\delta = M^{-1} \text{ and } \psi(t) = c(t) - f.x(t)\ .$$

C. Derived PCC

According to the CNC principle, it is possible to enhance the performance of a PCC by enabling switching at high frequencies, i.e. $t_e \to 0^+$.

To understand the effect of fast switching, let's rewrite the equation of the closed loop structure in the case of an output feedback:

$$y_{k+1} = C.f.x_k + C.f.\int_0^{t_e} e^{-A.\tau}.B.\gamma.e^{\alpha.\tau}d\tau.\lambda_k \quad (15)$$

with $\dim(\lambda(t)) = m$, $\dim(\alpha) = m \times m$, $\dim(\gamma) = r \times m$.

By realizing $t_e \to 0^+$, it is possible to simplify (15) by

$$y_{k+1} = \Upsilon.x_k + \Upsilon.\int_0^{t_e} (I_n - A.\tau).B.\gamma.(I_n + \alpha.t_e).d\tau.\lambda_k \quad (16)$$

with $\Upsilon = C.(I_n + A.t_e)$.

We can thus write (16) as

$$y_{k+1} = y_k + C.A.t_e.x_k + (C.B.\gamma.t_e + \varepsilon(t_e^{\,2})).\lambda_k \quad (17)$$

$\varepsilon(t_e^{\,2})$ being negligible when $t_e \to 0^+$.

In order to evaluate the initial condition of the PCC state at each switching instant, we fix the tracking condition as $y_{k+1} = c_{o,k}$, where $c_o(t)$ is the output's desired trajectory, such that the closed loop structure becomes

$$(C.B.\gamma.t_e + \varepsilon(t_e^{\,2})).\lambda_k = c_{o,k} - y_k - C.A.t_e.x_k \quad (18)$$

In order to solve (18) numerically, we rewrite the latter as

$$\lambda_k - [\lambda_k - (C.B.\gamma.t_e + \varepsilon(t_e^{\,2})).\lambda_k] = c_{o,k} - y_k - C.A.t_e.x_k \quad (19)$$

with fast switching ($t_e \to 0^+$), equation (19) becomes

$$\lambda_k = I_m^-.\lambda_k + c_{o,k} - y_k \quad (20)$$

Equation (20) can be interpreted algorithmically by an iterative evaluation of λ_k at each calculation step:

$$\lambda_k \leftarrow I_m^-.\lambda_k + c_{o,k} - y_k \quad (21)$$

The calculation of the initial condition of the PCC state at each switching instant is thus highly simplified.

Moreover, the structure of the PCC is simplified by the fact that $t_e \to 0^+$. In fact, in this condition, the evolution of the controller's state is negligible, such that the integrator setup of figure 2a acts as a Zero Order Holder (ZOH). Furthermore, if we consider that switching occur at each calculation step of a digital calculator, the ZOH can be replaced by a short circuit illustrated as Fig. 3. Thus the major advantage of this adaptation of the usual PCC is the fact that the model of the plant is unnecessary.

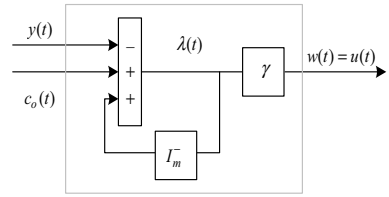
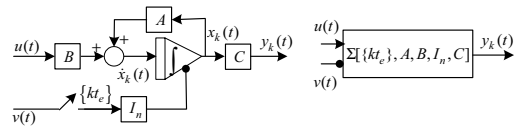


Fig. 3. Derived Piecewise Continuous Controller (DPCC)

IV. Piecewise Continuous Observer

The rested problem is how to estimate the x-y robot's x-axis output position $x_C(t)$ with the sampled and delayed θ_{k-1}. The presented PCO which consists mainly on a RODLO and two PCS systems responds to this demand.

In reference to section 2 and equation (3-4), a PCS $\Sigma(\{kt_e\},$ *A, B, I_n, C*) in the Φ_k piece, which is illustrated by Fig. 4, can be defined as follows:



a) PCS detailed structure b) PCS symbolic representation

Fig. 4. Piecewise Continuous Observer

$$x_k = v(kt_e),\ k \in S$$

$$x_k(t) = \exp(A(t - kt_e))x_k + \int_{kt_e}^{t} \exp A(t-\tau)Bu(\tau)d\tau, t \in \Im - S$$

$$y_k(t) = Cx_k(t),\ t \in \Im$$

Thus, the PCO is constructed as the following three steps and its realization structure is presented in Fig. 5.

A. First step

Using the PCS I, with *u(t)* and *v(t)=0* :

$$M_{k-1}(t) = \int_{(k-1)t_e}^{t} \exp A(t-\tau)Bu(\tau)d\tau,$$

By sampling (ZOH) at each kt_e, one obtains:

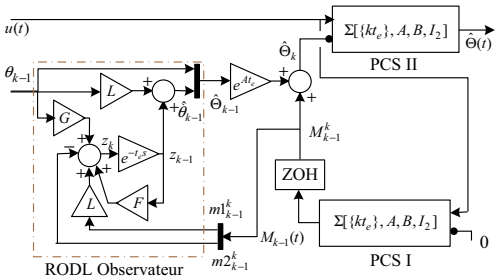


Fig. 5. Piecewise Continuous Observer

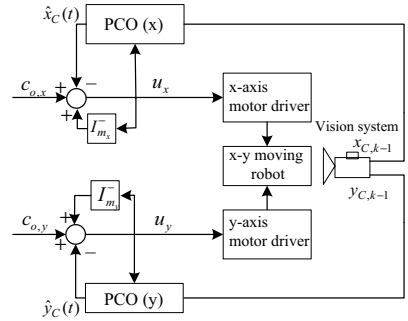


Fig. 6. Tracking control architecture for vision based x-y robot

$$M_{k-1}^{k}=\int_{(k-1)t_e}^{kt_e}\exp A(kt_e-\tau)Bu(\tau)d\tau=\begin{bmatrix} m1_{k-1}^{k} \\ m2_{k-1}^{k}\end{bmatrix}.$$

B. Second step

Θ_{k-1} is estimated by a RODL observer defined below:

$$z_k = Fz_{k-1}+G\theta_{k-1}+(m2_{k-1}^{k}-Lm1_{k-1}^{k})$$

$$z_{k-1}=\hat{\dot{\theta}}_{k-1}-L\theta_{k-1}$$

where F, G and L are defined from $\exp(At_e)=\begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22}\end{bmatrix}$ as:

$F=(f_{22}-Lf_{12})$,

$G=(f_{22}-Lf_{12})L+(f_{21}-Lf_{11})$,

$L=(f_{22}/f_{12})\in R$ (imposed value to maximize the RODL Observer's convergence speed)

Estimating $\dot{\theta}_{k-1}$ by $\hat{\dot{\theta}}_{k-1}=z_{k-1}+L\theta_{k-1}$, one gets $\hat{\Theta}_{k-1}$, then $\hat{\Theta}_k$, by integration of (14) on the time interval Φ_{k-1}:

$$\hat{\Theta}_k=\exp(At_e)\hat{\Theta}_{k-1}+M_{k-1}^{k}.$$

C. Third step

Using the PCS II, with the same input *u(t)* and $v(t)=\hat{\Theta}_k$:

$$\hat{\Theta}(t)=\exp(At)\hat{\Theta}_k+\int_{kt_e}^{t}\exp A(t-\tau)Bv(\tau)d\tau.$$

Thus one has the estimated position: $\hat{x}_C(t)=\hat{\theta}(t)=C.\hat{\Theta}(t)$.

V. Experimental Results

A. Trajectory control architecture

In view of validating the presented method, one has tested in the real vision based x-y robot, the whole tracking control architecture shown in Fig. 6.

The desired output's trajectories for each axe x and y are: $c_{o,x}$ & $c_{o,y}$. And it's important to note that, comparing to the structure of DPCC in Fig. 3, one chooses $\gamma_x=1$, $\gamma_y=1$, and the values of $I_{m_x}^{-}$ and $I_{m_y}^{-}$ are imposed to 0.9.

B. Experimental Results

The desired tracking trajectories for each axis are defined as follows:

$$\begin{bmatrix} c_{o,x} \\ c_{o,y}\end{bmatrix}=\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha\end{bmatrix}\begin{bmatrix} m\sin\omega_x t \\ n\cos\omega_y t\end{bmatrix},$$

where α is the oblique angle, *m* & *n* are amplitudes and ω_x & ω_y are the imposed different frequencies for each axis.

One tested the chosen values as follows: α=120°, $m=4$ (cm), $n=4$ (cm), $\omega_x=2.5$ (rad/s), $\omega_y=7.5$ (rad/s) and its corresponding results are illustrated in Fig. 7-10.

It's important to remind that $x_C(t)$ & $y_C(t)$ are continuous x-y robot's position measured by encoders, $\hat{x}_C(t)$ & $\hat{y}_C(t)$ are estimated x-y robot's position by PCO and $x_{C,k-1}$ & $y_{C,k-1}$ are sampled and delayed measurements by camera.

From Fig. 7-9, it's easily to contact that, the presented PCO compensate very well the camera's sampled and delayed feedback. The tracking performance of the presented DPCC is particularly illustrated in Fig. 9. And for comparison sake, one plots the $c_{o,x}$ & $c_{o,y}$ and $x_C(t)$ & $y_C(t)$ in the same x-y plane. And the corresponding tracking control u_x & u_y are shown in the Fig. 10.

In order to test the robustness of the tracking method, one covers the x-y robot's camera measurements for an instant and the corresponding result is illustrated in Fig. 11. The saturated peak in Fig. 11(a) is produced by covering the camera's feedback and considered as a perturbation for the PCO and DPCC.

From the preferred results one concludes that the presented method is effective and robust even if the camera drops out some measurements for an instant.

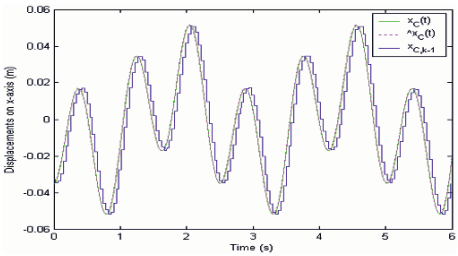


Fig. 7. Comparison of measurements on x axis

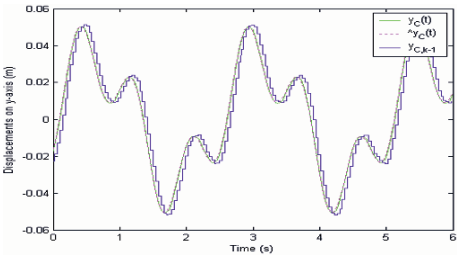


Fig. 8. Comparison of the measurements on y axis

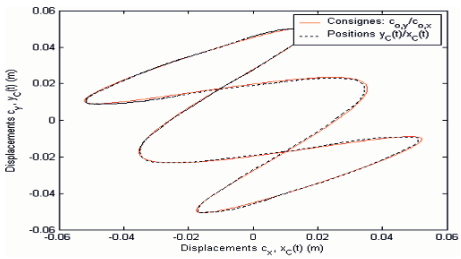


Fig. 9. Tracking result without perturbation

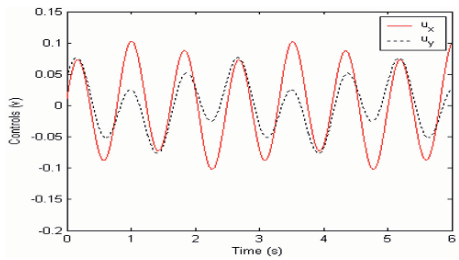
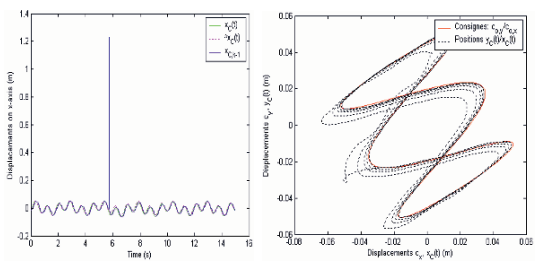


Fig. 10. Tracking control without perturbation



(a) (b)

Fig. 11. Tracking result under perturbation

VI. Conclusion

The proposed PCC and PCO which is based on the PCS in this paper is appropriate for control of linear plants in cases where the only available feedback comes from a sensor (low-cost camera in our case) delivering the plant's output vector in a sampled and delayed format. The proposed DPCC for trajectory tracking is independent of the plant's model and is reliable and robust from the experimental tests. And it's important to note that the theory of PCS (especially the PCO) can be used to control a vision based inverted pendulum which is nonlinear and unstable, and the available videos are accessible on:
http://www-lagis.univ-lille1.fr/~wang/Research_eng.html.txt.

Acknowledgement

This work was supported by the Nord - Pas de Calais region, the France and the European Community with the contract 15010/02Y0064/03-04 CAR/Presage N° 4605 Obi. 2-2004:2 - 4.1 - N° 160/4605. The authors would like to thank them for their kind support.

References

[1] M. S. Branicky, V. S. Borkar, and S. K. Mitter, "A unified framework for hybrid control, In Proceedings IEEE Conf. Decision Contr., Lake Buena Vista, 1994, pp. 4248-4234.

[2] A. Chamroo, C. Vasseur, and H.P. Wang, "Plant control using digital sensors that introduce a delayed and sampled output," *11th ELMA World Congress* , Vol. 1, pp. 119-124, Sofia, Bulgaria, Sep. 2005.

[3] D.W. Deley, "Controlling an inverted pendulum: example of a digital feedback control system," Available: http://members.cox.net/srice1/pendulum/index.htm.

[4] J. Heikkila, and O. Silven, "A four-step camera calibration procedure with implicit image correction," *1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.1106-1112, Jun 1997.

[5] S. Hutchinson, G.D. Hager, and P.I. Corke, "A tutorial on visual servo control," *IEEE Trans. on Robotics and Automation*, Vol. 12, No. 5, pp. 651-670, Oct. 1996.

[6] V. Koncar, and C. Vasseur, "Control of linear systems using piecewise continuous systems," *IEE Control Theory & Applications*, Vol. 150, n° 6, pp. 565-576, Nov. 2003.

[7] Y. Koren, and O. Masory, "Reference-word circular interpolation for CNC system," *Trans. ASME J. Eng. Ind.*, 104, pp. 400–405, 1982.

[8] F. Laurent, Sur la commande d'un filtre linéaire par des impulsions multimodulées. *C.R. Acad. Sc, Paris*, t. 270, pp. 288-289.

[9] F.J. Lin, R.J. Wai, and P.K. Huang, "Two-axis motion control system using wavelet neural network for ultrasonic motor drives, *IEE Proc.-Electr. Power Appl.*, Vol. 151, No. 5, pp. 613-621, Sep. 2004.

[10] M. Tittus, and B. Egart, "Control Design for Integrator Hybrid Systems," IEEE Transaction on Automatic Control, Vol.43, No.4, 1998, pp. 491-500.

[11] C. Vasseur, Contribution à l'étude des systèmes échantillonnés commandés par impulsions multimodulées, *Ph.D. Thesis*, University Lille 1, Villeneuve d'Ascq, France, 1972.

[12] H.P. Wang, C. Vasseur, A. Chamroo, and V. Koncar, "Sampled tracking for delayed systems using piecewise functioning controller," *4th IEEE conf. on Computational Engineering in Systems Applications*, Vol. 2, Beijing, China, Oct. 2006.

[13] H.R. Wang, X. Wang, and C.D. Jiang (2004). Adaptive fuzzy friction compensation for X-Y Table," *Proc. of the 23rd Chinese Control Conference,* pp. 1165-1168

Reliability Model for MEMS Accelerometers

Xingguo Xiong[1], Yu-Liang Wu[2], Wen-Ben Jone[3],
[1] Department of Electrical and Computer Engineering, University of Bridgeport, Bridgeport, CT 06604, USA
[2] Dept. of Computer Science & Engineering, The Chinese University of Hong Kong, Shattin, Hong Kong
[3] Department of ECECS, University of Cincinnati, Cincinnati, OH 45221, USA
Email: [1] xxiong@bridgeport.edu, [2] ylw@cse.cuhk.edu.hk, [3] wjone@ececs.uc.edu

Abstract— **MEMS (Microelectromechanical System) reliability is a very critical issue for its commercial applications. In order to measure the reliability of MEMS, a systematic reliability model is required. In this paper, we developed a MEMS reliability model for quantitative assessment of the MEMS reliability analysis. Based on this model, we analyze the reliability of both BISR (built-in-self-repairable) and non-BISR MEMS comb accelerometers under Z-axis shocking environment. Simulation results demonstrate very effective reliability enhancement due to the BISR design. The reliability model can also be applied to other MEMS devices under various failure mechanisms in a similar way.**

Keywords: MEMS (Microelectromechanical System), redundancy repair, reliability, accelerometer, fracture probability.

I. INTRODUCTION

In order for MEMS technologies to be used for real applications, yield as well as reliability are two very important issues which need to be immediately addressed. Reliability analysis is required for almost every commercial product. With the commercialization of MEMS devices, their reliabilities need to be thoroughly studied. Second, MEMS will be integrated into System-on-Chip (SoC) design very soon. The reliability of an entire SoC cannot be guaranteed if the reliability of MEMS is low. Further, MEMS is finding more and more applications in safety-critical areas, such as aerospace, medical instruments. For these applications, extremely high reliability is required. For example, during the launching process of a rocket, the failure of a tiny MEMS device may easily lead to unpredictable disaster. An unreliable bioMEMS chip embedded inside human body can be a serious threat to the health and may lead to the loss of life. Thus, the reliability research for MEMS in safety-critical applications is an especially urgent need.

Efforts on MEMS failure mechanisms and reliability research have been reported [1]-[6]. In [1], a comprehensive analysis on various MEMS defect sources, failure mechanisms and how they affect the MEMS reliability have been proposed. In [2][3], the MEMS material fatigue and aging under long-term cyclic loading has been discussed. In [4][5], MEMS reliability under shock and vibration environments has been explored. In [6], efforts have been made to explore the physical mechanism of stiction in surface micromachining and its impact on MEMS reliability. Although its importance has been widely recognized by researchers, MEMS reliability research still remains in its infant stage. Multiple energy domains are generally involved in the working principle of MEMS devices. Moreover, most MEMS devices contain movable components. Hence, MEMS devices are vulnerable to much more defect sources during its fabrication process and in-field usage compared to VLSI chips. This makes MEMS reliability research a challenging work. The understanding of various MEMS failure mechanisms is also non-trivial. Many MEMS failure mechanisms are unclear yet. The interaction among different MEMS failure mechanisms and its influence on MEMS reliability are not well discussed. Furthermore, most of the MEMS reliability analysis methods are still qualitative. A well-developed quantitative MEMS reliability model is not available. In this work, we develop a MEMS reliability model which can be used to quantitatively evaluate the reliability of MEMS devices. The reliability model has been applied to quantitatively assess the reliability of our previously proposed self-repairable MEMS accelerometer [7]. Based on this model, we evaluate the reliabilities of both BISR (built-in self-repair) and non-BISR MEMS accelerometers under shock environments. A comparison of the simulation results demonstrates that the BISR design leads to an effective reliability enhancement when compared to the non-BISR device. Although here we analyze the MEMS reliability under shock environments as an example, the reliability model and the research strategies can also be applied to other MEMS devices under other failure mechanisms (e.g., material fatigue, stiction) in a similar way.

II. RELIABILITY ANALYSIS OF NON-BIST/BISR MEMS ACCELEROMETERS

MEMS comb accelerometer device contains a large number of finger groups which are necessary to ensure enough signal strength. However, the large number of finger groups unavoidably leads to the decrease in yield as well as reliability. In [7], we proposed a built-in self-repairable MEMS comb accelerometer device. The device consists of

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 261–266.

six identical modules (for example), and each module has its own beams, mass and finger structures (fixed and movable). By assumption, four modules are connected together as the main device, while the remaining two modules serve as redundancy. The movable parts of each module are physically connected to those of adjacent modules through the common anchors, and signals sensed by all movable fingers in the device are connected to the sensing circuit directly. However, the fixed fingers of each module are connected to the modulation signal circuit through switches made of analog MUXes. By turning on or off these switches, we can determine whether a module works as part of the main device or the redundant device. If a module is tested as faulty, the control circuit will permanently exclude the module from the main device and replace it with a good redundant module (if there is any). Thus, after repairing, the main device can still be ensured to work properly. Because each module has its own independent beam and mass structure, a faulty module does not affect the function of other modules. For example, even if the movable part of one module is broken or stuck to substrate, the movable parts of other modules can still move freely and work jointly to ensure the function of the main device. Our previous analysis in [7] demonstrates that the BISR design leads to effective yield increase. In this paper, we will develop a reliability model and evaluate the effectiveness of reliability enhancement due to the BISR design. In the following discussion, we call the MEMS comb accelerometer with (without) the BISR feature as a BISR (non-BISR) accelerometer. In order for a fair comparison, we assume that the total number of finger groups of the non-BISR accelerometer should equal to that of the main device in the BISR accelerometer.

A. Basic Concepts of Reliability

Reliability is the probability that a component, equipment, or system will perform the required function under different operating conditions encountered for a stated period of time [8]. The reliability function is denoted by $R(t)$ $(0 \le R(t) \le 1)$ where t is time. The larger the reliability function value is, the more reliable the component, equipment or system will be. Meanwhile, the unreliability $Q(t)$, or the probability of failure, is defined as the probability that a component, equipment, or system will not perform the required function under the operating conditions encountered for a stated period of time t. Apparently, we have $R(t)+Q(t) = 1$. The failure rate λ is expressed as the ratio of the total number of failures to the total operating time: $\lambda = K/T$, where K is the number of failures and T is the total operating time. For most products, λ is generally a very small number. Its unit can be number of failures per 1×10^6 hours. As we can see, the larger the λ value is, the more unreliable the component, equipment or system will be.

The Mean-Time-To-Failures (MTTF) is the reciprocal of the failure rate: $MTTF = 1/\lambda$.

In reality, the failure rate λ is generally the function of time. The failure rate $\lambda(t)$ of mechanical components, VLSI chips and MEMS devices follows the behavior of a *bathtub* curve as shown in Figure 1 [8]. It is believed the failure rate of MEMS devices also follows the bathtub curve. The bathtub function consists of three regions. In the initial stage, the failure rate is high due to the latent defects in the device, and falls off till time t_{infat}. After that, the device enters a stable stage with a constant failure rate. For high reliability applications, this constant failure rate λ should be extremely small. Finally, after time $t_{operation}$, the failure rate will increase sharply due to wear-out and the device comes to the end of its lifetime. The useful time of the device with low constant failure rate is defined as $t_{useful} = t_{operation} - t_{infant.}$ Generally, $t_{infatand}$ and $t_{operation}$ of a certain product cannot be given as definite values. The manufacturers will give average values for their products based on statistical testing.

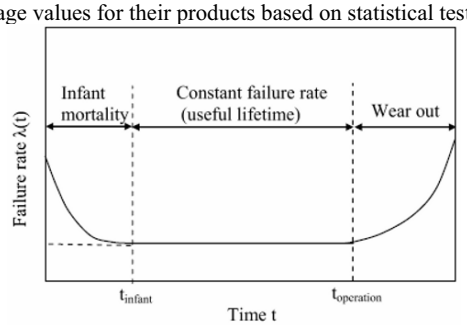


Figure 1. The bathtub curve of failure rate [8]

There are several standard probability models available for describing the reliability of a system: the *exponential* reliability distribution, the *binomial* reliability distribution, the *Poisson* reliability distribution, and *Weibull* reliability distribution [8].

1) The exponential reliability distribution

The exponential reliability distribution is the most common probability model used to predict the lifetime of a system. Thus, in our MEMS reliability analysis, the exponential reliability model is used. The reliability function $R(t)$ is expressed as

$$R(t) = e^{-\lambda t}$$

where λ is the failure rate and t is the time period. The unreliability $Q(t)$ is therefore

$$Q(t) = 1 - R(t) = 1 - e^{-\lambda t}$$

The exponential reliability distribution is shown in Figure 2.

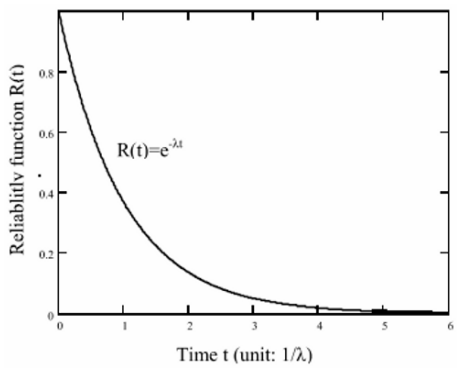


Figure 2. The exponential reliability distribution function.

2) The binomial reliability distribution

The binomial reliability distribution is used for describing the reliability of a discrete distribution. It can be expressed as $(R+Q)^n=1$, where n is the total number of trials conducted. A typical example of binominal reliability distribution is the case of flipping coins for heads and tails. If we define the probability of obtaining a head and a tail as R and Q separately, we have $R=Q=0.5$. If two trials are made (n=2), then $(R+Q)^2=R^2+2RQ+Q^2=1$.

3) The Poisson reliability distribution

The Poisson reliability distribution is a discrete distribution which provides a useful tool in the case of the binomial distribution. The reliability can be expressed as

$R=1-(Q_1+Q_2+Q_3+\ldots)$

where Q_i is the probability of exactly i failures occurring during time period t, and can be expressed as

$Q_i(t)=(\lambda t)^i e^{-\lambda t}/i!$

The probability of zero failure in the Poisson reliability model comes to the result of the exponential reliability distribution $R_0(t)=e^{-\lambda t}$. That is, the exponential reliability distribution can be treated as a special case of Poisson reliability model for $i=0$.

4) The Weibull reliability distribution

The Weibull reliability function is expressed as

$R(T1)=exp\{-[(T1-\gamma)/\eta]^{\beta}\}$

For general reliability measurement, we consider $\gamma=0$ and $\alpha=1/\eta$, hence

$R(t)=exp[-(\alpha t)^{\beta}]$

where β is the shape parameter which indicates whether the failure rate is increasing or decreasing. If $\beta<1.0$, the failure rate is decreasing. If $\beta=1.0$, the failure rate is constant. If $\beta>1.0$, the failure rate is increasing. If $\beta=1.0$, it comes to the result of exponential distribution: $R(t)=e^{-(\alpha t)}$. Thus, the exponential distribution can also be treated as a special case of the Weibull reliability function for $\beta=1.0$.

The above reliability models describe the reliability of an individual component. However, sometimes we also need to evaluate the reliability of a system consisting of multiple components. Such system reliability models enable us to calculate the reliability characteristics of a design, evaluate design alternatives, and perform sensitivity analysis. Depending on the configuration of the system, different system reliability models (such as series, parallel, and k-out-of-n, etc.) are available [8].

1) Series Reliability Model

If the functional operation of a system depends on the successful operation of all system components, the reliability of the system can be calculated with a series reliability model. The reliability block diagram of the series model is shown in Figure 3. Assume the system consists of nnumber of serial components, and the failure rate of each component i is λ_i, the reliability of the entire series system is

$$R_{tot}=\prod_{i=1}^{n}R_i(t)=e^{-\lambda_1 t}\cdot e^{-\lambda_2 t}\ldots\cdot e^{-\lambda_n t}=e^{-\left(\sum_{i=1}^{n}\lambda_i\right)t}$$

From the above equation, we have

$$\lambda_{tot}=\sum_{i=1}^{n}\lambda_i$$

That is, the failure rate of a series system is the sum of the failure rates of all the series components. The mean-time-to-failure (MTTF) of the system can be expressed as

$$MTTF_{tot}=\sum_{i=1}^{n}\frac{1}{MTTF_i}$$

where $MTTF_{tot}$ is the MTTF of the entire series system, $MTTF_i$ is the MTTF of the i^{th} component.

Figure 3. The block diagram of series reliability model

2) Parallel Reliability Model

The block diagram of a parallel reliability model is shown in Figure 4. The system functions properly until all of the components (C_1 to C_n) fail. Hence, the reliability R_{tot} of the parallel system is given as one minus the probability of failure for each component as shown below

$$R_{tot}=1-\prod_{i=1}^{n}\left[1-R_i(t)\right]=1-\prod_{i=1}^{n}\left[1-e^{-\lambda_i t}\right]$$

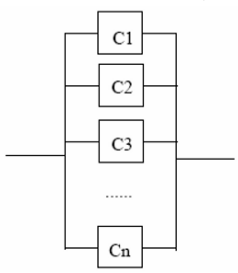


Figure 4. The block diagram of parallel reliability model.

3) k-out-of-n Redundancy Reliability Model

The block diagram of a k-out-of-n redundancy reliability model is shown in Figure 5. Among the n number of modulus, at least k number of modules need to be fault-free in order for the whole system to work properly. In other words, a maximum of n-k number of faulty modules are allowed without losing the function of the whole system. Our BISR comb accelerometer [7] is exactly an example of k-out-of-n reliability model. If the BISR accelerometer contains n modules in the main device and m redundant modules, it is a n-out-of-(n+m) model. Assume the reliability of each component as R_c. The reliability R_{tot} of the k-out-of-n redundancy system is given as below [8]

$$R_{tot}(t) = R_c^n(t) + nR_c^{n-1}(t)(Q_c(t)) + \ldots + \binom{n}{k}(R_c(t))^k(Q_c(t))^{n-k}$$

Assume the failure rate of each component as λ, the MTTF of the system can be calculated as [8]

$$MTTF = \frac{1}{\lambda}\left(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \ldots + \frac{1}{k}\right)$$

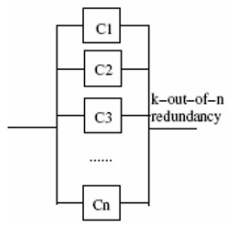


Figure 5. The block diagram of k-out-of-n redundancy reliability model.

B. Reliability Model of Non-BISR MEMS Device

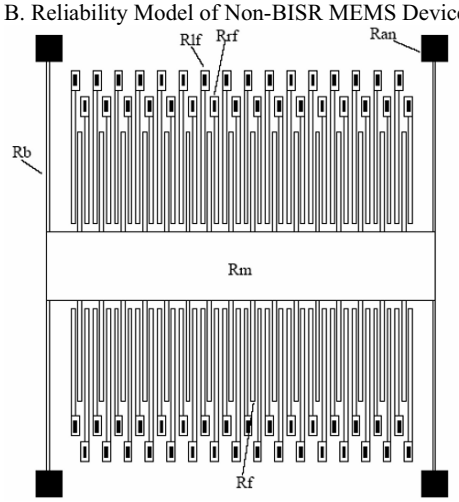


Figure 6. The structural diagram of a non-BISR MEMS device.

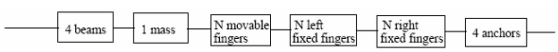


Figure 7. The reliability model for the non-BISR MEMS device.

The structural diagram of the non-BISR MEMS accelerometer [7][9] for this analysis is shown in Figure 6, and its corresponding reliability model is shown in Figure 7. As shown in Figure 6, the non-BISR device consists of four beams, one mass, N number of movable fingers, $2N$ number of left/right fixed fingers, and four anchors. All these components must be fault-free in order for the entire device to function correctly. If any of these components becomes faulty, the whole device will malfunction. For example, if one of the beams is broken, a movable finger is stuck to the corresponding fixed finger, or the mass is stuck to the substrate, then the device will not be able to work properly. Although the fixed components such as anchors and fixed fingers are less likely to be faulty than the movable components, they also must be fault-free in order to ensure the proper function of the device. For example, if one left fixed finger is shorted to its neighborhood right fixed finger, this will lead to a short circuit to the signal detection circuit. Or, if an anchor is lifted off the substrate, the device structure will be unstable. Thus, the reliability of a non-BISR device can be described as a series model, as shown in Figure 7. Since the four beams have exactly the same dimension and equal loading, we assume all four beams have the same reliability R_b. Similarly, we assume the reliabilities of the mass and each movable finger as R_m and R_f separately. The reliabilities of one fixed finger and one anchor are denoted as R_{ff} and R_{an} individually. Finally, the reliability of the entire non-BISR device can be expressed as

$$R_{nsr} = R_b^4 \cdot R_m \cdot R_f^N \cdot R_{ff}^{2N} \cdot R_{an}$$

In reality, movable components of the MEMS device are vulnerable to more defect sources when compared with fixed components. Thus, the reliabilities of fixed components can be much higher than those of movable components. The defects of fixed components mainly result from device fabrication. The devices with catastrophic defects in fixed components can be filtered out during manufacturing test. During in-field usage, it is less likely for the fixed components to develop new defects compared to the movable components. Thus, in the following discussion, we will mainly concentrate on the reliability of the movable components, and the reliability of the fixed parts can be treated as constant 1. In this way, the reliability of the non-BISR device can be expressed as

$$R_{nsr} = R_b^4 \cdot R_m \cdot R_f^N$$

From the reliability model of the non-BISR device, we can see that the major threat to the device reliability comes from the large index (i.e., N) of R_f. For example, if the reliability of a single movable finger is 0.99 and N equals 42, this will reduce the entire device reliability to 0.656 even though the beam and mass are assumed perfect. This is the major impetus for us to implement the redundancy repair technique for MEMS comb accelerometers. By

modularizing the device, each module contains a smaller number of comb fingers. The reliability of each module will be higher than the original non-BISR device. By implementing the redundancy repair technique, even higher reliability can be achieved for the entire device. Assume the failure rates for the beam, mass and finger as λ_b, λ_m and λ_f separately. The reliability functions of the beam, mass and finger can be expressed as

$$R_b = e^{-\lambda_b t}, R_m = e^{-\lambda_m t}, R_f = e^{-\lambda_f t}.$$

Hence, the reliability function of the non-BISR device with respect to time t is

$$R_{nsr}(t) = e^{-\lambda_{nsr} t} = e^{-(4\lambda_b + \lambda_m + N\lambda_f)t}$$

where the failure rate λ_{nsr} of the non-BISR device is

$\lambda_{nsr} = 4\lambda_b + \lambda_m + N\lambda_f$

Once we know the failure rates λ_b, λ_m and λ_f, we can derive the reliability of the non-BISR device at certain time t based on the equation for $R_{nsr}(t)$.

C. Reliability Model of BISR MEMS Device

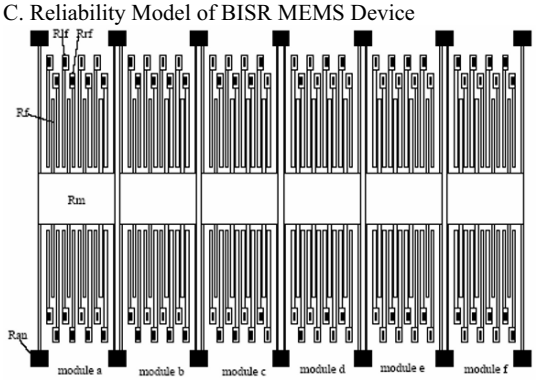


Figure 8. The structural diagram of the BISR MEMS accelerometer.

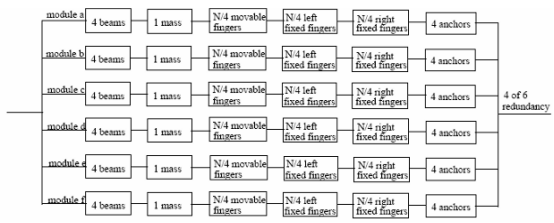


Figure 9. Reliability model for BISR device.

The structural diagram of the BISR comb accelerometer [7] is shown in Figure 8 and its corresponding reliability model is shown in Figure 9. The BISR comb accelerometer consists of six identical modules. In each module, it consists of four beams, one smaller mass, (N/4) number of movable fingers, (N/2) number of left and right fixed fingers, and four anchors. Note that N is the number of movable fingers for the non-BISR device. Again, the reliability of each module can be described with a series model. Assume the reliabilities of each beam, movable finger and mass as R_{br}, R_{fr} and R_{mr} separately, and the reliabilities of each fixed finger and anchor as R_{ff} and R_{an} individually. The reliability R_{mod} of each BISR module can be expressed as

$$R_{\mathrm{mod}} = R_{br}^4 \cdot R_{mr} \cdot R_f^{N/4} \cdot R_{ff}^{N/2} \cdot R_{an}$$

Similarly, if we only concentrate on the reliability of movable components, the reliability of each module can be expressed as

$$R_{\mathrm{mod}} = R_{br}^4 \cdot R_{mr} \cdot R_f^{N/4}$$

Assume the failure rates for the beam, mass and finger of each BISR module as λ_{br}, λ_{mr} and λ_f separately. The reliability functions of the beam, mass and finger can be expressed as

$$R_{br} = e^{-\lambda_{br} t}, R_m = e^{-\lambda_{mr} t}, R_f = e^{-\lambda_{fr} t}$$

Hence, the reliability function of the entire BISR module is

$$R_{\mathrm{mod}}(t) = e^{-(\lambda_{\mathrm{mod}} t)} = e^{-[4\lambda_{br} + \lambda_{mr} + (N/4)\lambda_f]t}$$

where the failure rate λ_{mod} of each module is

$\lambda_{mod} = 4\lambda_{br} + \lambda_{mr} + (N/4)\lambda_f$

The unreliability Q_{mod} of each module is

$Q_{mod}(t) = 1 - R_{mod}(t)$

There are totally six identical modules in the BISR comb accelerometer. Among them, only four modules are required to be fault-free to ensure the proper function of the device. The reliability of such a system can be calculated using a k-out-of-n redundancy reliability model. The reliability R_{sr} of the BISR accelerometer can be expressed as

$$R_{sr}(t) = R_{\mathrm{mod}}^6(t) + 6R_{\mathrm{mod}}^5(t)Q_{\mathrm{mod}}(t) + 15R_{\mathrm{mod}}^4(t)Q_{\mathrm{mod}}^2(t)$$

$$= \left(e^{-\lambda_{\mathrm{mod}}(t)}\right)^6 + 6\left(e^{-\lambda_{\mathrm{mod}}(t)}\right)^5\left(1 - e^{-\lambda_{\mathrm{mod}}(t)}\right) + 15\left(e^{-\lambda_{\mathrm{mod}}(t)}\right)^4\left(1 - e^{-\lambda_{\mathrm{mod}}(t)}\right)^2$$

This is the reliability function of a 4-out-of-6 redundancy BISR comb accelerometer. For the general case, assume there are n number of modules in the main device and m number of modules as redundancy. The reliability of the BISR accelerometer can be expressed as

$$R_{sr}(t) = \left(R_{\mathrm{mod}}(t)\right)^{n+m} + (n+m)\left(R_{\mathrm{mod}}(t)\right)^{n+m-1}\left(Q_{\mathrm{mod}}(t)\right) + \ldots$$

$$+ \binom{n+m}{n}\left(R_{\mathrm{mod}}(t)\right)^n\left(Q_{\mathrm{mod}}(t)\right)^m$$

D. Reliability Enhancement and Reliability Analysis

After we have developed the reliability models for both non-BISR and BISR MEMS devices, we can derive the reliability increase $IR(t)$ due to redundancy repair by

$IR(t) = R_{sr}(t) - R_{nsr}(t)$

In the BISR MEMS comb accelerometer, the large number of comb finger groups are divided into several modules. Each module contains a smaller number of comb finger groups, and thus the risk for each module to be faulty is lowered. The redundancy further improves the reliability of the BISR device. However, compared to the original non-BISR device, the BISR device contains more beams. This will lead to the decrease of the reliability. The net reliability

increase or decrease depends on the interaction between these counteracting factors.

III. Reliability Analysis Result

Based on the above reliability model, we can simulate the reliability increase due to redundancy repair for BISR MEMS accelerometers compared to non-BISR design. For example, we evaluated the reliabilities of example designs for both non-BISR and BISR devices under shock environment, and assume fracture as the major failure mechanism. The non-BISR device has a beam width of $W_{bnsr} = 3.2\mu m$. For BISR device, we consider two cases [10]:
(1). BWC device: the BISR device with Beam Width Compensation (i.e., shrinking the beam width to compensate the sensitivity loss due to BISR design) with
$W_{bbwc} = 2.0\mu m = 0.63 W_{bnsr}$.
(2). EFC device: the BISR device with Electrostatic Force Compensation (i.e., using electrostatic force to compensate the sensitivity loss) with $W_{befc} = 3.2\mu m = W_{bnsr}$.

We used ANSYS simulation to extract the maximum stress σ_{max} values corresponding to about 30 different Z-axis shock accelerations from $0g$ to $10^5 g$. Based on these data, we can calculate the Weibull fracture probability (P_f) [11] and the shock reliability (R_f) values for non-BISR and BWC/EFC BISR devices corresponding to each individual shock acceleration. The plots of the simulated Z-axis shock reliability versus each shock acceleration of non-BISR/BISR devices are shown in Figure 10. As shown in the Figure 10, the shock reliability of the EFC BISR device is larger than that of the BWC BISR device, and both are larger than the non-BISR device. We can see that if the Z-shock acceleration falls into the range between the mean fracture stress of non-BISR device and BISR devices, a (maximum) reliability increase as large as 1 can be achieved due to the BISR design.

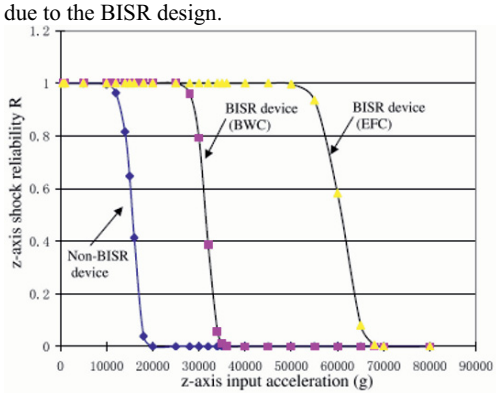


Figure 10. The z-axis shock reliability for non-BISR/BISR devices.

IV. CONCLUSIONS

In this paper, a quantitative MEMS reliability model has been developed. Based on this reliability model, the reliabilities of both non-BISR and BISR MEMS comb accelerometer devices under shock environment have been assessed. The Weibull fracture probability function has been used to evaluate the fracture probability of MEMS accelerometers under Z-axis shock acceleration. The simulation results demonstrate an effective reliability enhancement for both the BWC and EFC BISR comb accelerometers, compared to non-BISR design. In the future research, we will also try to further improve our MEMS reliability model to account for other MEMS devices under various failure mechanisms. In this way, the reliability model can be used to assess and predict the MEMS reliability for various MEMS devices under various failure mechanisms. It can also guide us in improving the MEMS reliability for safety-critical applications.

REFERENCES

[1] B. Stark (editor), "MEMS Reliability Assurance Guidelines for Space Applications", Jet Propulsion Laboratory Publication 99-1, Pasadena, USA, Jan. 1999.

[2] M. Tabib-Azar, K. Wong, andW. Ko, "Aging Phenomena in heavily doped (p+) micromachined silicon cantilever beams", *Sensors and Actuators A*, Vol. 33, pp. 199-206, 1992.

[3] C. L. Muhlstein, R. T. Howe, and R. O. Ritchie, "Fatigue of polycrystalline silicon for microelectromechanical system applications: crack growth and stability under resonant loading conditions", *Mechanics of Materials*, Vol. 36, pp. 13-33, 2004.

[4] D. M. Tanner, J. A.Walraven, K. Helgesen, L.W. Irwin, F. Brown, N. F. Smith, and N. Masers, "MEMS reliability in shock environments", *Proceedings of IEEE International Reliability Physics Symposium*, San Jose, CA, USA, pp. 129-138, Apr. 10-13, 2000.

[5] D. M. Tanner, J. A. Walraven, K.S.Helgesen, L. W. Irwin, D. L. Gregory, J. R. Stake, and N. F. Smith, "MEMS reliability in a vibration environment", *IEEE 38th Annual International Reliability Physics Symposium*, San Jose, California, USA, pp. 139-145, 2000.

[6] W. M. V. Spengen, R. Puers, and I. D. Wolf, "On the physics of stiction and its impact on the reliability of microstructures", *Journal of Adhesion Science and Technology*, Vol. 17, No. 4, pp. 563-582, 2003.

[7] X. Xiong, Y. Wu, and W. Jone, "Design and analysis of self-repairable MEMS accelerometer," *Proceedings of the 20th IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems (DFT'05)*, Monterey, CA, USA, pp. 21-29, Oct. 3-5, 2005.

[8] K. Neubeck, Practical Reliability Analysis, Pearson Prentice Hall, New Jersey, USA, 2004.

[9] W. Kuehnel, and S. Sherman, "A Surface Micromachined Silicon Accelerometer with On-chip Detection Circuitry," *Sensors and Actuators*, Vol. A, Issue 45, pp. 7-16, 1994.

[10] X. Xiong, Built-in self-test and self-repair for capacitive MEMS devices, Ph.D dissertation, University of Cincinnati, July 2005.

[11] S. Greek, F. Ericson, S. Johansson, and J. Schweitz, "In situ tensile strength measurement and Weibull analysis of thick film and thin film micromachined polysilicon structure", *Thin Solid Films*, Vol. 292, pp. 247-254, 1997.

Diagram, Dynamic Geometry and *Sangaku*

Yoshiteru Ishida[1,2] and Masayuki Fujisawa[1]
[1]Department of Knowledge-Based Information Engineering,
Toyohashi University of Technology,
Tempaku, Toyohashi, 441-8580 Japan
[2]Intelligent Sensing System Research Center,
Toyohashi University of Technology,
Tempaku, Toyohashi, 441-8580 Japan

Abstract **We have developed a web-based system of presenting *Sangaku*. *Sangaku* is a way of presenting theorems and conjectures in geometry in one sheet of votive picture. *Sangaku* was used for temples in Japan to exchange geometrical findings, and has been studied as a traditional form of Japanese mathematics called *Wasan*. We developed a diagrammatic reasoning that directly uses geometric objects as primitives for reasoning, while tools called dynamic geometry software have also been developed. In developing the web-based *Sangaku* presentation system, we used diagrammatic reasoning which allows users to directly manipulate geometric objects. As a result, geometric findings can be presented not in a static manner of a single picture, but in a dynamic manner of operational geometric objects. Some examples of *Sangaku* can be seen at: http://www.sys.tutkie.tut.ac.jp/MuJapan.html.**

I. INTRODUCTION

Diagrammatic reasoning has been proposed [1] as a form of human reasoning based on diagrams with the aid of visual information processing. One possible explanation for the power of diagrammatic reasoning may be the human internal image [2], which is a geometrically operational model. As an external realization of the internal image model, we have developed a diagrammatic reasoning that allows geometrical objects to be directly manipulated as primitives for reasoning [3].

On the other hand, there are many dynamic geometry software tools ranging from free software to commercial ones. Dynamic geometry software allows us to draw composite geometrical objects that can be operated dynamically by using the *ruler and compass method* [4]. As the name of this method suggests, some auxiliary objects are required, so it would be difficult to make the procedure independent of the target composite objects. To solve this problem we implemented the diagrammatic reasoning by Java™. The system allows us to draw a composite geometric object in a constraint-based and component-oriented way. To demonstrate this, we have used the system to express Japanese *Sangaku* (a sheet of votive picture expressing a geometric finding), and so we named this system which expresses *Sangaku* dynamically the "*Sangaku* system". The goal of the tool is to demonstrate the potential of diagrammatic reasoning for E-learning and other domains requiring operation-based understanding of composite geometric objects.

The *Sangaku* system is designed to make users (drawers of *Sangaku*) to focus on the constraint between two primitive objects. Before explaining the details of our *Sangaku* system, we must first explain *Sangaku*: a Japanese traditional geometry.

II. *SANGAKU*

Sangaku is a votive picture expressing a geometric finding painted on one sheet [5]. Traditionally, *Sangaku* were dedicated to temples during the Edo period in Japan, to give thanks to God for solving the geometric problem. There are about 900 *Sangaku* in existence. Originally, *Sangaku* were drawn with both the geometric problem and the answer together, but later only the geometric problem was drawn and dedicated, and then persons who found the answer to the problem dedicated their *Sangaku* separately.

Temples used to be places where many people gathered, and people of all classes had a common interest in geometry. The competition to dedicate *Sangaku* led to advances in Japanese traditional geometry and excellent ones can be found. Figure 1 shows an example of a geometric composite object drawn by the *Sangaku* system. It asserts the geometric finding that the four centers ***F, G, H*** and ***I*** of the four circles ***C1, C2, C3*** and

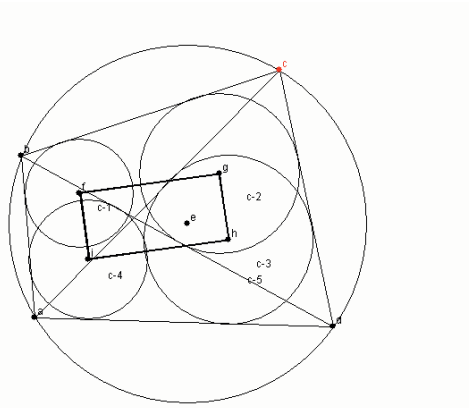


Figure 1. An example of SANGAKU stating that the four centers ***F, G, H*** and ***I*** of the four circles ***C1, C2, C3*** and ***C4***, each of which is tangential to a diagonal and two adjacent sides of a quadrilateral ***ABCD*** touching internally with a circle, form a rectangle ***FGHI***. See Figure 18 also.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 267–272.

C4 always form a rectangle ***FGHI*** whenever the following constraint is satisfied: each circle (e.g. ***C1***) is inscribed to the triangle (e.g. ***ABC***) formed by adjacent lines and a diagonal line of the quadrilateral ***ABCD***.

III. WEB-BASED SYSTEM FOR PRESENTING *SANGAKU*

We developed *Sangaku,* a web-based system to present *Sangaku* in a dynamic fashion, using a Java™ applet. The user can visually understand the geometric finding by directly manipulating the object and observing the continuous transformation of the object. The data of a *Sangaku* is saved in a server by Java™ servlet (Figure 2). Some examples of *Sangaku* can be seen at:
http://www.sys.tutkie.tut.ac.jp/MuJapan.html.

Figures 16, 17 and 18 in the appendix show some examples of *Sangaku* presented in the *Sangaku* system. Since continuous transformation cannot be presented on paper, we have presented the *Sangaku* before and after the continuous transformation.

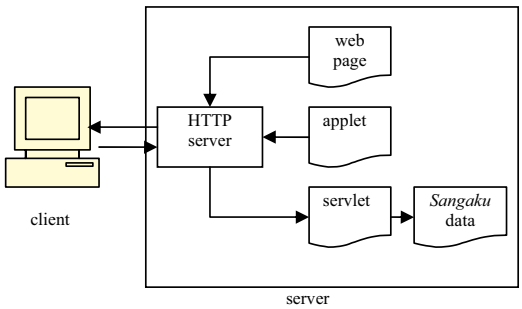
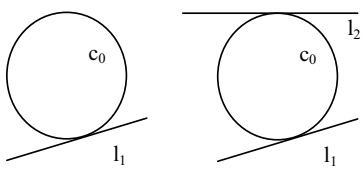


Figure 2. A web-based system of presenting *Sangaku* based on JAVA servlet.

IV. CONSTRAINT HANDLING IN *SANGAKU*

One feature of the *Sangaku* system is that it allows users to construct a composite diagram by incrementally adding constraints. These constraints can be defined in a "declarative construction", and hence permit intuitively recognizable and reusable constraints definition and construction. For example, consider a diagram of two geometric objects: a line l_1 and a circle c_0 satisfying the constraint that the line must always be tangential to the circle as shown in Figure 3 (a).



(a) Diagram A (b) Diagram B

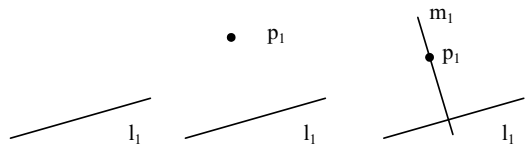
Figure 3 Examples of composite diagrams. (a) a line l_1 tangential to a circle c_0, (b) another line l_2 is also tangential to the circle c_0.

A. *Construction of diagrams*

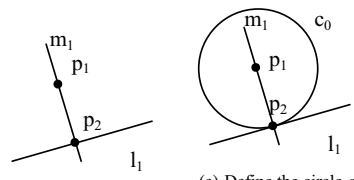
First, a procedure for constructing Diagram A (a composite geometric object) can be stated as follows:

1. *Define the line l_1. (primitive element)*
2. *Define the point p_1. (sub-object)*
3. *Define the line m_1 through the point p_1 that is perpendicular to the line l_1. (sub-object)*
4. *Define the point p_2 that is the intersection point of the lines m_1 and l_1. (sub-object)*
5. *Define the circle c_0 through the point p_2 the center of which circle is the point p_1. (primitive element)*

The primitive elements of p_1, p_2 and m_1 are sub-objects. This process shows that the "procedural construction" cannot define the circle c_0 without defining the line l_1 (Figure 4).



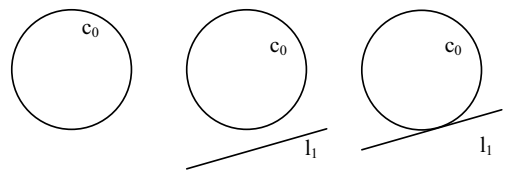
(a) Define the line l_1.(b) Define the point p_1. (c) Define the m_1 line perpendicular to l_1.



(d) Define the point p_2.

(e) Define the circle c_0.

Figure 4. Procedural construction.



(a) Define the circle c_0. (b) Define the line l_1. (c) Declare the constraint relation.

Figure 5. Declarative construction.

Secondly, the "declarative construction" follows.

1. *Define the circle c_0. (primitive element)*
2. *Define the line l_1. (primitive element)*
3. *Declare that the line l_1 is tangential to the circle c_0. (a constraint)*

This process shows that "declarative construction" requires fewer and simpler operations (Figure 5).

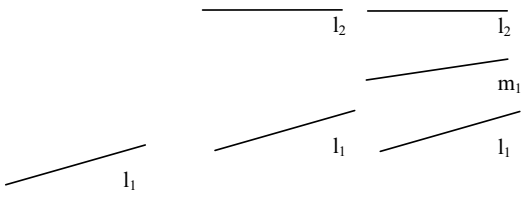
B. Addition of constraint

We stated that the *Sangaku* system allows incremental addition of constraints. Here, we compare the procedural and declarative construction with respect to constraint addition using the example of Figure 3.

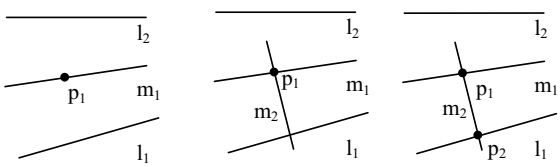
First, the procedure for constructing Diagram B by adding the constraint to Diagram A is as follows:

1. *Define the line* $\boldsymbol{l_1}$*. (primitive element)*
2. *Define the line* $\boldsymbol{l_2}$*. (primitive element)*
3. *Define the line* $\boldsymbol{m_1}$ *that is the bisector of the angle between the lines* $\boldsymbol{l_1}$ *and* $\boldsymbol{l_2}$*. (sub-element)*
4. *Define the point* $\boldsymbol{p_1}$ *on the line* $\boldsymbol{m_1}$*. (sub-element)*
5. *Define the line* $\boldsymbol{m_2}$ *through the point* $\boldsymbol{p_1}$ *that is perpendicular to the line* $\boldsymbol{l_1}$*. (sub-element)*
6. *Define the point* $\boldsymbol{p_2}$ *that is the intersection point of the lines* $\boldsymbol{m_2}$ *and* $\boldsymbol{l_1}$*. (sub-element)*
7. *Define the circle* $\boldsymbol{c_0}$ *through the point* $\boldsymbol{p_2}$ *the center of which circle is the point* $\boldsymbol{p_1}$*. (primitive element)*

This process shows that the procedural construction does not allow Diagram A to be reused (Figure 6).

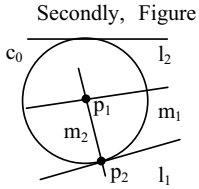


(a) Define the line $\boldsymbol{l_1}$. (b) Define the line l_2. (c) Define the line m_1.



(d) Define the point $\boldsymbol{p_1}$. (e) Define the perpendicular to $\boldsymbol{m_2}$. (f) Define the point $\boldsymbol{p_2}$.

Secondly, Figure 7 shows a declarative construction of



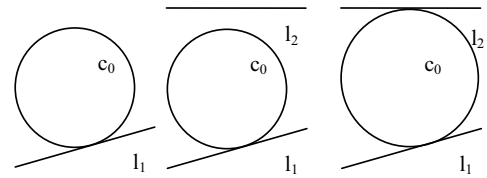
(g) Define the circle $\boldsymbol{c_0}$.

Figure 6. Procedural addition of a constraint

Diagram B by adding the constraint to Diagram A as follows:

1. *Prepare Diagram A. (reuse diagram)*
2. *Define the line* $\boldsymbol{l_2}$*. (primitive element)*
3. *Declare that the line* $\boldsymbol{l_2}$ *is a tangent of the circle* $\boldsymbol{c_0}$*. (constraint relation)*

In contrast to the procedural addition of the constraint, the declarative addition of the constraint allows Diagram A to be reused. Further, the primitive elements are defined independently from the constraint added (Figure 7).



(a) Prepare Diagram A. (b) Define the line $\boldsymbol{l_2}$. (c) Declare constraint.

Figure 7. Declarative addition of constraint.

V. THE *SANGAKU* SYSTEM

A. Primitive elements and constraints

Primitive elements prepared in the *Sangaku* system are as follows:

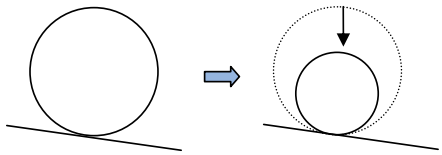
- Point
- Line
- Circle

Constraints in the *Sangaku* system are shown in Table 1. By defining a constraint between the above primitive elements, a composite object of the diagram can be constructed.

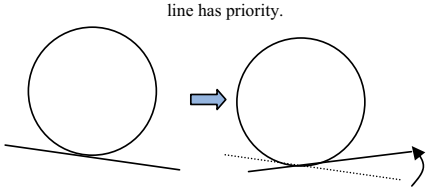
TABLE 1 Constraints in the *Sangaku* system

Constraint	Explanation
LineHandle	Define the end-point of a segment of a line.
CircleHandle	Define the center-point of a circle.
PointOnLine	Define the point on a segment of a line.
PointOnCircle	Define a point on a circle.
LineLength	Stabilize a segment of a line length.
Angle	Stabilize an angle.
TangentLineToCircle	Define a segment of a line tangent to a circle.
TangentCircleToCircle	Define a circle tangent to a circle.

The *Sangaku* system allows an object to be transformed, which results in dynamically propagating constraints and transforming other objects so that the constraints must always be satisfied. To do that, the *Sangaku* system explicitly permits priority allocation among primitive elements. Figure 8 shows a comparison when priority is allocated to different elements.



(a) The continuous transformation when the line has priority.



(b) The continuous transformation when the circle has priority.

Figure 8. Priority allocation between two primitive elements in *Sangaku*

B. *Example of diagram construction by the Sangaku system*

Figure 9 shows a screenshot of drawing diagrams in the *Sangaku* system. Drawing can be done by mouse on the drawing area. The element drawn on the drawing area can be moved by dragging and dropping it.

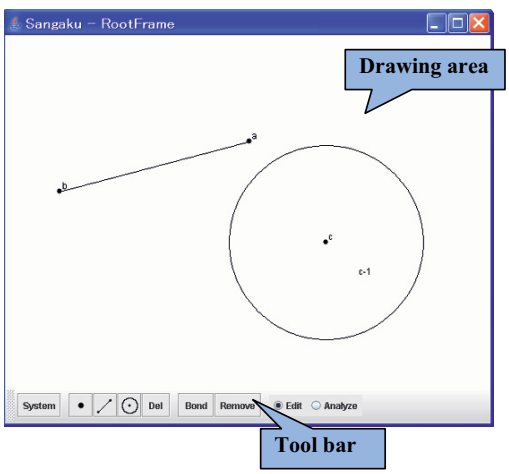
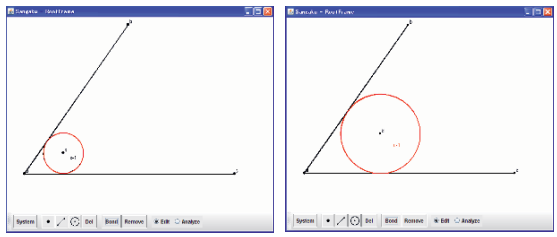


Figure 9. A screenshot of drawing diagrams in *Sangaku.*

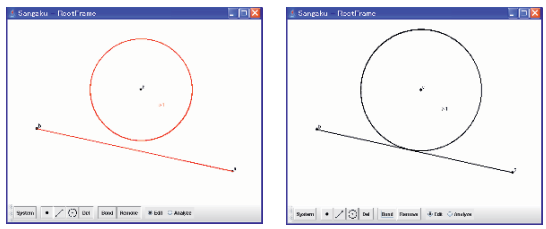
Figure 10 shows two snapshots: before and after the diagram is continuously transformed. Two lines are tangential to a circle, and this constraint is satisfied during the continuous transformation of the circle.

Figure 11 shows an example of declaring a constraint between two primitive objects.



(a) Before the transformation. (b) After the transformation.

Figure 10. Example of continuous transformation.



(a) Select the line and the circle. (b) Declare the constraint and the diagram is transformed

Figure 11. Constraint addition to two elements.

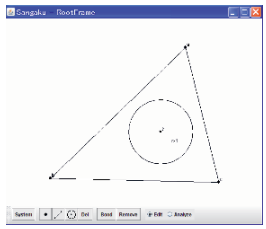
Figure 12 is an example of drawing the inscribed circle of a triangle in the *Sangaku* system. As shown in Figures 3 and 5, the compass and ruler method requires the bisector of an angle and other sub-objects to be drawn; however, the *Sangaku* system requires only four steps.

VI. Constraint Propagation in the *Sangaku* System

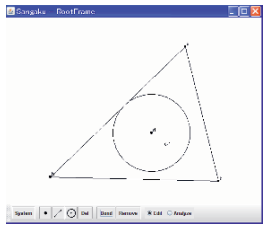
In the *Sangaku* system, any composite objects are composed of primitive elements and constraints among the elements. The *Sangaku* system has two basic classes: the Diagram class and the Constraint class. Each class corresponds to primitive elements and constraints, respectively. A composite object may be constructed as a network structure that a Constraint class combines two Diagram classes as shown in Figure 13.

A transformation of one object will trigger the constraint propagation and the propagated constraint again causes another object to transform. In this manner, the constraint and transformation will be propagated. Figure 14 shows an example of constraint propagation. The diagram has a constraint that the point $\boldsymbol{p_1}$ is the end of the lines $\boldsymbol{l_1}$ and $\boldsymbol{l_2}$. When the point $\boldsymbol{p_1}$ is moved, the line $\boldsymbol{l_2}$ will be transformed accordingly to satisfy the constraint. Then, the line $\boldsymbol{l_1}$ again transforms to satisfy the constraint. In an actual *Sangaku*

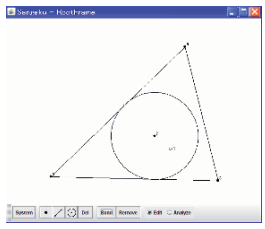
session, these processes will not be drawn but only the transformed diagram will be depicted.



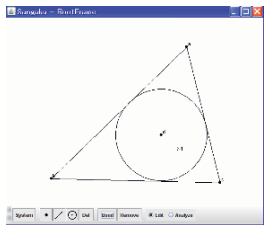
(a) Draw a circle and a triangle.



(b) Declare the constraint.



(c) Declare the constraint.



(d) Declare the constraint.

Figure 12. The procedure to draw inscribed circle of a triangle.

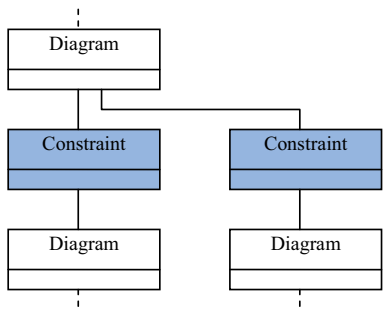


Figure 13. Constraint structure.

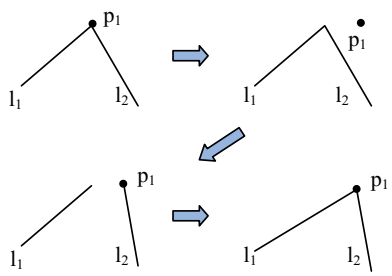


Figure 14. An example of constraint propagation.

Although the basic constraint structure is a tree as shown in Figure 13, it would have a returning link to the upper level as well in a complex composite object; hence the structure can be a network. In that case, the constraint is propagated back and forth. This is equivalent to solving simultaneous equations, and indeed there is a dynamic geometry software tool that performs constraint propagation by solving such equations [6]. However, the *Sangaku* system uses a different method.

Figure 15 shows a data flow in the *Sangaku* system. The *Sangaku* system is implemented by connection between an applet and a servlet. The data of a composite object drawn by the *Sangaku* system (as shown in Figure 13) is passed to the servlet. The XML document is generated by relating the primitive elements with a constraint, and the servlet is saved as a file.

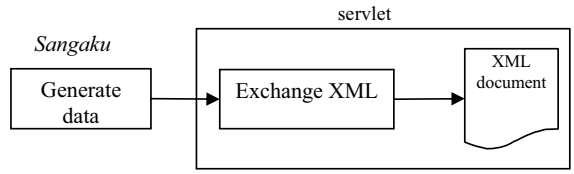


Figure 15. The data flow in *Sangaku*.

The XML document has the following entries and attributes, and defines the structure of the diagram (a composite object).

First, the entry and attribute that define the primitive element are as follows. A primitive element and information regarding coordinates etc. are defined.

Secondly, a constraint is imposed as shown below. A constraint and two primitive elements restricted by the constraint are defined.

- primitive element

```
<diagram id="l1" type="line">
<x1>650.0</x1>
<y1>182.0</y1>
<x2>700.0</x2>
<y2>300.0</y2>
<display>yes</display>
<hold>yes</hold>
</diagram>
```

- constraint

```
<constraint type="handle">
<diagram ref="l1" type="line"/>
<diagram ref="a" type="point"/>
</constraint>
```

VII. Conclusion

We have developed a diagrammatic reasoning system and applied it to present *Sangaku*, which is a diagrammatic picture indicating a geometrical finding in one sheet. The web-based *Sangaku* presentation system is implemented with a "declarative construction" of *Sangaku* (a composite object).

The objective of the *Sangaku* system is to assist studying elementary geometry in a constructive and creative manner, as a method of E-learning geometry.

ACKNOWLEDGMENTS

This work was supported by the Global COE Program "Frontiers of Intelligent Sensing", from the Ministry of Education, Culture, Sports, Science and Technology. This work was also supported in part by Grants-in-Aid from Toyohashi University of Technology.

REFERENCES

[1] J. H. Larkin and H. A. Simon, "Why a diagram is (sometimes) worth a thousand words," Cognitive Science, Vol. 11, 1987.
[2] S. M. Kosslyn, Image and Mind, Harvard University Press, 1980.
[3] Y. Ishida, "Diagrammatic Reasoning by Propagating Constraints through Geometric Objects," *The Ninth International Workshop on Qualitative Reasoning about Physical Systems* (QR95), pp. 256-263, 1995.
[4] N. Jakiw, *Geometer's Sketchpad, User Guide and Reference Manual*, Key Curriculum Press, Berkeley 1992.
[5] H. Fukagawa and D. Pedoe, *Japanese Temple Geometry Problems*, Morikita Publishing, 1991 (in Japanese).
[6] X. S. Gao and Q. Lin, "MMP/Geometer – A Software Package for Automated Geometric Reasoning," *Automated Deduction in Geometry* 2002, pp. 44-66, 2002.

APPENDIX: EXAMPLES OF *SANGAKU*

Some examples of *Sangaku* drawn by the tool are shown in Figures 16, 17 and 18. There are many circle-related problems, and when combined, increase the complexity.

Figure 16 is an example of a *Sangaku* stating a geometric relation among four circles. The radius of the nth circle R_n can be expressed by the radii of two large circles R_1 and R_2 as:

$$R_n = \left(\frac{n-2}{\sqrt{R_1}} + \frac{1}{\sqrt{R_2}} \right)^2$$

Figure 17 is an example of a *Sangaku* on two circles and a triangle. Two circles are tangential to an equilateral triangle. The sum of the radii of these circles R_1+R_2 is always constant.

Figure 18 (also in Figure 1) is an example of a *Sangaku* stating a geometric relation among five circles. A quadrilateral is inscribed in a circle. The quadrilateral is divided into four triangles by the two diagonal lines. There are four inscribed circles of the triangle. A tetragon formed by the lines connecting the centers of the circles always forms a rectangle.

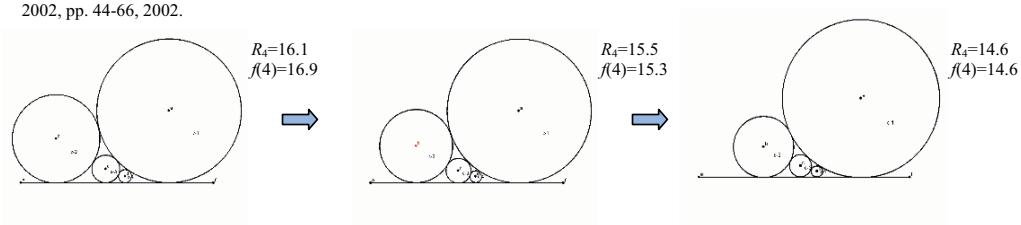


Figure 16. *Sangaku* stating a geometric relation among four circles.

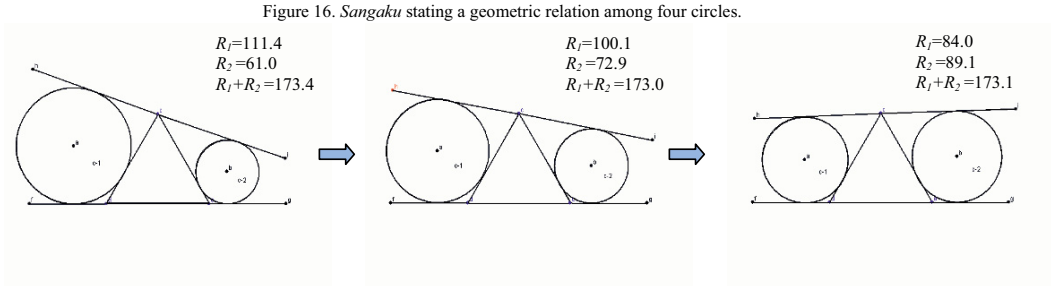


Figure 17. *Sangaku* stating a geometric relation among two circles and a triangle.

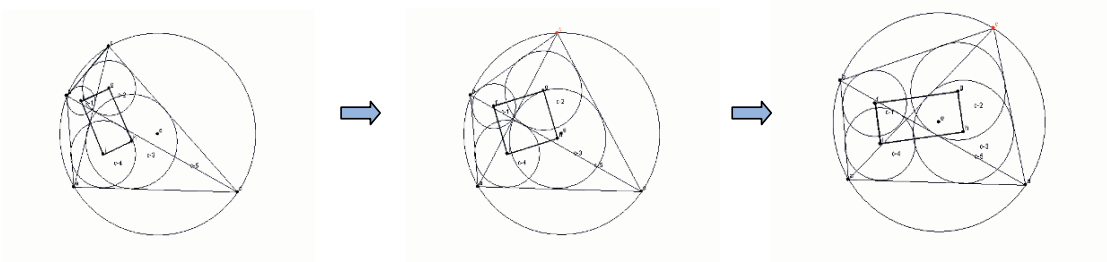


Figure 18. *Sangaku* stating a geometric relation among five circles. See Figure 1 also.

A Modeling Technique for Execution and Simulation of Discrete Automation

Yuval Cohen*

Department of Management and Economics
The Open University of Israel
Raanana, Israel, 43107

Ming-En Wang, Bopaya Bidanda

Department of Industrial Engineering
The University of Pittsburgh
Pittsburgh, USA, 15261

Abstract *- This paper presents a technique for automatic generation of a transition network and its translation to a PLC program from the processing specifications. On the one hand the paper presents an algorithm for generating PLC code, and on the other hand the paper presents a new representation scheme expressing the high level function of the system. Thus, it is hoped that the new transition networks could become successful replacement for the current PLC languages.*

I. Introduction

Major part of discrete industrial automation hardware is controlled by Programmable Logic Controllers (PLCs) [1,2]. These controllers are programmed using one or more of the five standard low-level PLC programming languages, defined by the IEC-1131-3 standard (the most popular of the five being Ladder Diagram (LD).[3]) The ever increasing need for better modelling and implementation of automation was not satisfactorily fulfilled by any of the existing techniques [4]. There is ongoing research for a simpler, faster, and friendlier model, that would be easier to change and debug [5].
This paper is an extension of [6, 7] which are more general. This paper presents the details of automating the PLC code generation and modification. The purpose of this technique is to enable the functional specification to generate almost automatically a high level graphical model that can be translated to a PLC program and implemented on currently available shop-floor controllers.
The proposed approach seeks the following three targets:
1. To automate the generation of high-level graphical model as much as possible.
2. To facilitate understanding of the automation logic and process by programmers and operators (which may maintain, change or debug the model).
3. To enable translation of the graphical model into PLC code, for controlling switches, actuators, and sensors.

The proposed high-level graphical tool describing the process and its logic is a special version of a state-transition network. A care was paid to stay compatible with the ongoing development of the XTND transition-net standard of the world wide web consortium (http://www.w3.org/TR/xtnd/).
The rest of the paper continues as follows: section II presents the modeling scheme. Section III discusses the automatic generation of the transitions from minimal specifications. The first part of the paper presents and illustrates the modeling technique while the second part deals with its translation to ladder diagram.

II. The graphical model

The graphical model itself has two levels of detail:

- Standard state diagram
- Transition details

The state diagram is the standard state machine diagram with some additions such as conditions for state termination and for invoking a transition.

A. Actions State Diagram (ASD)

State diagrams can grow rapidly with the system size [8]. Therefore, we limit the state diagram to relatively simple tasks (less than, say, 20 states). Each state in the ASD describes an

* Corresponding author: Tel.: (972) 9-778-1883; Fax: (972) 9-778-0668; E-mail: yuvalco@openu.ac.il

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 273–277.

action (a single combination of outputs). Note that inputs are ignored at this stage. This not only eliminates the complexity of input-output relationships, but also provides a clearer view of a system's functionality and enables the designer to focus on small portions of information at a time. Since we tend to think of any discrete process in terms of actions, ASD offers a natural, simplified, and clear functional description. An ASD for an automatic drill press is depicted Figure 1. States are denoted by numbers and transitions by capital letters. The ASD does not include all the details regarding inputs, outputs, and variables. These details are embedded in E-Transitions, and discussed in sub-section B.

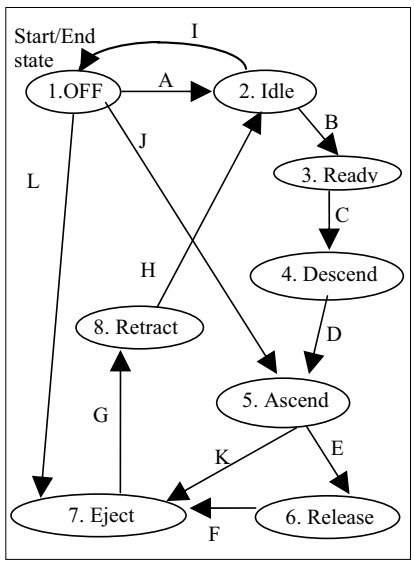


Fig. 1. An example of the Actions State Diagram — ASD

B. E-Transitions

The ASD is further exploded into a new type of a graphical scheme named E-Transition (for Elementary transition). E-Transitions describe the changes in low level elements such as inputs, outputs, and registers, required for executing the ASD. E-Transitions arrange the elements in a meaningful way that enables immediate comprehension of a low level code. The E-Transitions are composed of the following elements: 1) places, 2) triggers, and 3) arcs. These elements are all depicted in figure 1. Each transition is activated by one or more triggers. The triggers are denoted by triangles pointing at a thick vertical line that symbolizes the transition. Places (denoted by circles) represent the inputs, outputs, events, and variables.

Events are assigned places with additional symbol to denotes the type of event (turn ON, and shut OFF). Places that use non-binary data (e.g., timers and counters) are denoted by rectangles. Additionally, places are added for logically denoting the states of the system. For example, transition C from state 3 to state 4 uses the corresponding ST3 and ST4 variables.

Two arc types used to activate triggers are as follows:

1. An enable arc (⋯▶) the triggers can fire only while the source place holds a token.
2. A disable arc (⋯●) the triggers can fire only while the source place is OFF.

Enable and disable arcs are drawn with dashed lines to denote that they do not activate or deactivate elements. Tokens are used to denote activated places. Two types of arcs used to identify the effects of a transition as follows:

1. Activate arc (⋯▶) turns ON the place when the corresponding E-Transition is activated.
2. Deactivate arc (—⊠) turns OFF the place when the corresponding E-Transition is activated.

Each trigger is invoked by places linked to the trigger by enable or disable arcs. Note the usage of the source state (STi) variable of the E-Transition to facilitate trigger's identification as one of the trigger's conditions. After the trigger is activated, a transition from the source state (i) to another state (j) occurs immediately. Each E-Transition also resets the source state variable (STi) and sets the destination state variable (STj). Note that each trigger has only on E-Transition, but a transition may have more than one trigger. Finally, the E-Transitions can be integrated into the ASD as shown in figure 2.

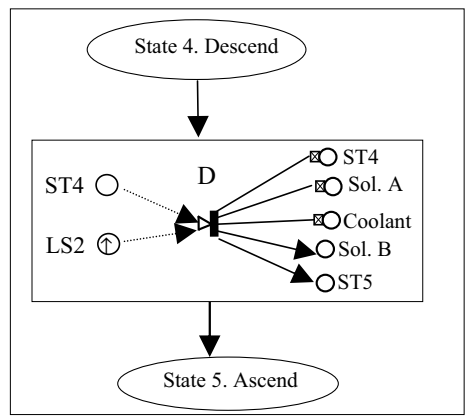


Fig. 2. A segment integrating an E-Transition for transition C in the ASD of figure 1.

III. E-TRANSITIONS GENERATION FROM MINIMAL DATA

This section presents a methodology for generating the E-Transitions using as little data as possible, while automating as many stages as possible.

While the ASD is relatively easy to construct, the details of the E-Transitions are much more detailed and even a slight error in specification may seriously affect the process. Therefore, E-Transitions generation should be done meticulously and preferably automated as much as possible.

The main input required for the proposed scheme, is a specification of the output and timers status for each ASD state. An example for such input (for a drill-press machine) is depicted in Table 1. This specification depends heavily on the specific system at hand, and must be done carefully and correctly.

Table 1. Status of outputs for each state of the ASD in Fig. 1 (entries associated with a mechanical actions or active timers are shaded.)

	OUTPUTS							
	Solenoids							
ASD State	A	B	C	D	Pump motor	Drill motor	Cool ant	Timer 1
1	off	off	off	off	off	off	off	off
2	off	off	off	off	on	off	off	on
3	off	off	on	off	on	on	off	off
4	on	off	on	off	on	on	on	off
5	off	on	on	off	on	on	off	off
6	off	off	off	off	on	off	off	off
7	off	off	off	on	on	off	off	Off
8	off	off	off	off	on	off	off	Off

Each transition in the ASD has origin state and destination state. The status changes between the origin and destination outputs and timers are the transition effects. Therefore, by comparing the lines in Table 1 that relate to the origin and destination states, the effects could be generated automatically.

Transition effects are summarized in Table 2. For example, it is easy to see that the changes between lines 1 and 2 in table 1 are in the pump motor and timer. These changes appear in line one of Table 2.

Table 2. The effects of the ASD transitions of Fig. 1 (generated by comparing the origin and destination state lines in Table 1.)

Transition	**Its effect/s**			
A (1→2)	Pump motor ↑	Timer ↑		
B (2→3)	Drill motor ↑	Solenoid C ↑		
C (3→4)	Solenoid A ↑	Coolant ↑		
D (4→5)	Solenoid A ↓	Coolant ↓	Solenoid B ↑	
E (5→6)	Solenoid B ↓	Solenoid C ↓	Drill motor ↓	
F (6→7)	Solenoid D ↑			
G (7→8)	Solenoid D ↓			
H (8→2)	Timer ↑	Reset ↓	Done ↑	
I (2→1)	Pump motor ↓	PROCESS2↓	Done ↓	
J (1→5)	Pump motor ↑	Solenoid B ↑	Drill motor↑	Reset↑
K (5→7)	Solenoid B ↓	Solenoid D ↑	Drill motor ↓	
L (1→7)	Pump motor ↑	Solenoid D ↑	Reset ↑	

The less trivial part is generating the trigger/s for each E-Transition. Each trigger event must be identified efficiently and effectively. Since each trigger has one transition and each transition has one origin state. The origin state is a natural candidate for trigger identification. Triggers must also have some event captured by an input

To facilitate trigger generation, a state indicator relay is dedicated for each state. While moving from one state to another, the relay signifying the origin state is turned OFF, and the relay signifying the next state is turned ON. Thus, the state relay is ON when the corresponding state is active, and OFF when the system is in other states. Triggers' activation is done when input changes happen during a given state. So each trigger is implemented by detecting the relevant input changes while the origin-state relay is ON. Part of the transition changes will be to update the current state relay.

Finding triggers involves finding the events for terminating their origin state. One way to do this is to identify active outputs (in the origin state) and for each such output identify related inputs, such as limit switches (LS). In our example, the active outputs are already shaded in Table 1. The related inputs for each are listed in Table 3. These inputs may trigger the transitions emanating from the origin state when changing their status by an external event. If the input status does not change as a result of the transition, it should not be part of the trigger.

The rest of the triggers are commands such as "process" or "reset". Table 3 gives the relevant details for trigger identification of each transition.

Table 3. Trigger events via source states terminating events

Tran si-tion	Source State	Source State Shaded entries from Table 1	Related Inputs	Speci-fied Varia-bles	Triggering events at the source state
A	1 (OFF)			Process↑	Process2↑ (Reset=off)
J	1 (OFF)			Reset↑	Reset↑ (LS1=off)
L	1 (OFF)			Reset↑	Reset↑ (LS1=on)
B	2 (Idle)	Timer 1			Timer↓
I	2 (Idle)	Timer 1			Timer↓
C	3 (Ready)	Sol. C	LS4		LS4↑
D	4 (Descend)	Sol. A Sol. C Coolant	LS2 LS4 cont.		LS2↑
E	5 (Ascend)	Sol. B Sol. C	LS1 LS4 cont.		LS1↑
K	5 (Ascend)	Sol. B Sol. C	LS1 LS4 cont.		LS1↑
F	6 (Release)	Sol. C (off)	LS3		LS3↑
G	7 (Eject)	Sol. D	LS6		LS6↑
H	8 (Retract)	Sol. D (off)	LS5		LS5↑

However, some of these combinations may be identical. See for example transitions *J* and *L* in Table 3. Therefore, there is a need to differentiate between their identical triggers. For this purpose supplement conditions are added. For example, for the trigger of *J*, LS1 must be OFF, and for the trigger of *L* it must be ON. This is evident from the right columns of both Table 3 and Table 4.

Table 3 reveals that transitions B and I have identical triggers that should be differentiated artificially by internal relays.

While each transition must have a trigger, it may happen that more than one trigger activate the same transition (such a case is not included in the example). Therefore, the analysis should lead to a list of all possible triggers. Each trigger has to be uniquely defined by different combination of: (1) origin-state relay, (2) input changes, and (3) supplementary identifiers. Such list of triggers is depicted in Figure 4.

Table 4. Minimum Identifiers for Trigger Events

Trigger event	Transi tion	Source state: Var. (name)	Detected Inputs Change	Supplementary Identifiers
1. Start	A	ST1 (closed)	PROCESS2↑	
2. Reset ascend	J	ST1 (closed)	Reset↑	LS1/**OFF**
3. Reset ejection	L	ST1 (closed)	Reset↑	LS1/**ON**
4. Warm-up completed	B	ST2 (Idle)	Timer ↓	*Done*/**OFF** (Internal variable turned OFF by Transition I)
6. Cool Down completed	I	ST2. (Idle)	Timer ↓	*Done*/**ON** (Internal variable turned ON by Transition H)
7. Clamp	C	ST3. (Ready)	LS4↑	
8. Down	D	ST4. (Descend)	LS2↑	
9. Up	E	ST5. (Ascend)	LS1↑	Reset/**OFF**
10. Reset	K	ST5. (Ascend)	LS1↑	Reset/**ON**
11. Released	F	ST6. (Release)	LS3↑	
12. Ejected	G	ST7. (Eject)	LS6↑	
13. Retract	H	ST8. (Retract)	LS5↑	

IV. SYSTEMATIC LADDER DIAGRAM GENERATION

A Ladder Diagram (LD) is chosen to illustrate the implementation of the model. The generated LD rungs are arranged in three main blocks as follows: 1) events identification 2) transition triggers, and 3) transition effects. Backward translation is also possible (Cohen and Bidanda, 1997) but is not presented here. The construction of the above three blocks is presented next.

A. Events Identification

Inputs and outputs change their voltage level when turned ON or OFF. These changes are referred as rising or falling edges. The international standard IEC 1131-3 defines special LD contacts for detecting rising and falling edges. A rising edge corresponds to a place with "↓" and a falling edge to a place with "↑".

B. Transition Triggers

Each trigger activates one E-Transition. Each transition is assigned an internal variable in the LD. When the E-Transition is enabled that variable will be turned ON. In order to implement this logic, a set of rules is described as follows:

I. Each trigger forms an LD rung.

II. Each place (in E-Transition) that is input to a trigger forms a contact: (enable arc forms a normally open (NO) contact, and disable arc a normally closed (NC) contact.
III. The LD rung output is a variable that corresponds to the activated transition.
Figure 3 depicts a ladder diagram segment corresponding to the triggers of transition D. These variables are used in figure 4.

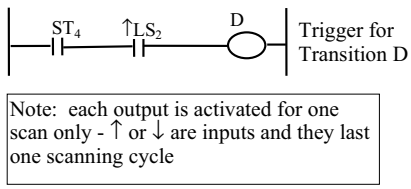


Fig. 3. Ladder Diagram segment for triggering transition D of the ASD in figures 1 and 2.

C. Transition Effects

The rules for establishing the ladder diagram portion of transition's effects is as follows:
1. Dedicate a rung for each output place of the E-Transition and add to it a corresponding LD output (e.g., the right hand places of figure 4 are translated into outputs in figure 6).
2. In each rung add a contact that corresponds to the relevant transition.
3. Activation arcs are translated into latched outputs, and Turn-off arcs are translated into Unlatched outputs.

Figure 4 depicts a ladder diagram segment corresponding to the effects of transition D

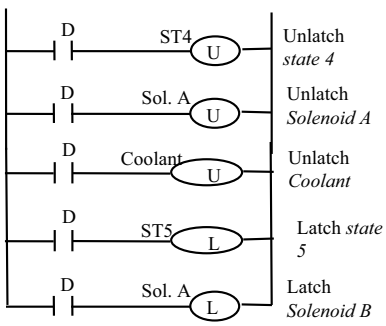


Fig. 4. A Ladder Diagram (LD) segment for the effects of transition D (see figure 2).

V. Conclusion

In this paper a discrete control modeling technique is presented that has a great potential for simplification of PLC code generation, verification, validation and maintenance. Such simplification could reduce the cost of discrete automation and could contribute to progress towards more flexible automation.

The proposed technique could be easily implemented using existing control equipment (PLCs). For each processing task, a simple state diagram that describes only the elementary actions of the system has to be built. From that point the process of the model generation could be automated, provided that the following input is specified: the status of outputs for each state; the related inputs of each state, and supplementary trigger identifiers.

Future research includes implementation of the proposed technique on an industrial shop floor. This would enable simulation and visualization of operation; it would enable high-level verification instead of code verification; it would assist in real-time tracking and failure analysis of the control system. Currently the construction of a translator (model to LD) is under way, to enable such future implementation.

References

[1] Collins K., PLC Programming for Industrial Automation, Exposure Publishing, 2007

[2] Bolton W., "Programmable Logic Controllers", Fourth Edition, Newnes, 2006

[3] Karl-Heinz J. and Tiegelkamp, M., IEC 61131-3: Programming Industrial Automation System, w. CD-ROM : Concepts and Programming Languages, Requirements for Programming Systems,6th edition, Springer, 2001.

[4] Brown A. B. and Hellerstein J. L., "Reducing the cost of IT operations—is automation always the answer?", In HotOS,, June 2005.

[5] Peng S. and Zhou M. C., "Sensor-based Petri net modeling for PLC stage programming of discrete-event control design", Proceedings of IEEE International Conference on Robotics and Automation (ICRA), Volume 2, 1907 - 1912, 2002.

[6] Cohen, Y. and Bidanda, B., "A discrete control modeling technique for automated industrial systems", Proceedings of the Embedded Computing Conference (ECC-96), Paris, France, pp. 279-287, 1996.

[7] Cohen, Y. and Bidanda, B., "A new discrete control modeling technique for automated industrial systems", Technical Report 97-2, Dept. of Industrial Engineering,University of Pittsburgh, 1997.

[8] Ferdinand Wagner (Author), Ruedi Schmuki (Author), Thomas Wagner (Author), Peter Wolstenholme (Author) Modeling Software with Finite State Machines: A Practical Approach, AUERBACH, 2006

Using DES in a Modified Design to Keep it from Oblivion

Abdelshakour Abuzneid, Florina Ochis, Rashmi Bajracharya, Lisa Rajbhandari, Shailaza Shrestha,

University of Bridgeport, CT

abuzneid@bridgeport.edu, eochis@bridgeport.edu, rbajrach@bridgeport.edu , lrajbhan@bridgeport.edu, shailazs@bridgeport.edu

Abstract:

DES, an old but very useful algorithm of encryption, has been endangered to be replaced in the latest years by the more advanced and competent algorithms as we can exemplify: Triple DES and AES (Advanced Encryption Standard) both from the same category – block oriented encryption algorithms. Obviously, the more advanced the technology is, the requirements for implementation are more costlier and often, the devices that support these advanced implementations are very sophisticated and hard to build or require a lot of work comparing with the classic, traditional once. From the financial as well as from implementation point of view, it would be definitely very important for the cryptography to keep using DES as long as possible, this being dictated of the ability of this encryption algorithm to remain efficient and how many resources might still have.

The present study is needed to be another way to look at DES and review the other side of improvements that might be done in order to keep it up for a few more years. This might be considered a hybrid between DES and Triple DES by keeping on one side the efficiency of plain DES and taking on the other side a part of Triple DES higher security. Even if is not a "Double DES" because of the different principle used, it might be a good idea attempting to mix these two important algorithms. For the purpose of this article we will call this implementation "2K-DES".

I. INTRODUCTION:

The advance in computer technology represents a widely known fact and from cryptography point of view it has a good part and a bad part. The good part is that encryption may take a huge advantage of the fast development of both hardware and software – old algorithms can be improved and new algorithms can be developed based on cumulative intelligence in the computer field.

The bad part might be considered the permanently increased vulnerability of the data in front of the cyberterrorism. In another words the technology explosion is taking place for "the good guys as well as for the bad guys".

One of the parents of the advanced encryption technology of today is DES (Data Encryption Standard) which "was a reasonably competent block cipher for its time, but advances in computing machinery have blown its security away" [12]. Still searching deeper into DES there might be a few more aspects that need to be taken advantage of.

II. D.E.S.

A. The Continuously Need For Improved Encryption Products

There is a known fact that if at the early beginning, the inventors of DES were estimated a 5 millions dollars brute - force machine [12] that would be able to crack it. The cruel reality showed that the predicted DES safety might be touched with much less financial expenses and the alarm signal was turned on for the first time in 1997 when a group of hobbyists broke DES [13]. Since than, a few more successful attempts have taken place and today a DES cracker machine is estimated to cost less than quarter million dollars. In this case the cryptographers needed to move ahead and use strong parts of the old algorithms they inherited for new once, more efficient and safer.

B. The Way in Which DES Encryption Has Touched Higher Performance Levels So Far

During the latest years, improving DES meant working both at the software and hardware level. Therefore the quality of the coding used was higher and higher and at the devices were more efficient, cheaper and easier to implement. In parallel with these structural and material improvements of the algorithm the power of the cracking machines has increased exponentially with the years passed since DES encryption algorithm has been developed.

From now on, it looks like no improvements over DES will be developed, especially due to the existence of other, more advanced encryptions procedures. However, something else can always be done…

C. Another Attempt to Keep DES "In Shape"

Beside all improvements mentioned before another attempt of improving DES could be interesting to analyze. The main idea is the following: As we already know, Data Encryption Standard is a block cipher algorithm meaning that it takes a plain text block of 64 bits as an input producing an output of the same number (64) of bits as a cipher text; the procedure being the same for the decryption, just in reverse order.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 278–282.

As many specialists concluded regarding DES weakness, Achilles heel is its short key length (56 bits).

However, the next step in DES hierarchy was represented by Triple DES and AES which both have a longer key (or can be considered keys) length: Triple DES, a modified version of its parent DES employing 112- or 168-bit keys and AES respectively, which operates with public-key, sizes of 128, 192, and 256 bits.

D. 2K-DES Description

In order to avoid the vulnerability induced by the DES's short key, another approach might be considered as following:

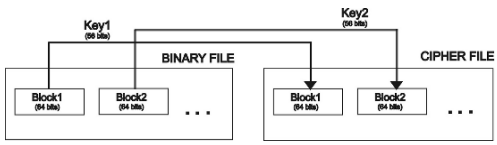


Fig. 1 *2K-DES process description*

The first input block of 64 bits – Block1 is encrypted with the first arbitrary generated key – Key1. The resultant Block1 is the first 64 bits encrypted binary cipher. Furthermore, the next 64 bits block will be taken encrypted with the second key - Key2 and the output represents the second block cipher., and so on until the end of the file is reached. Note that Key1 and Key2 are different, that means the **same text** is encrypted with **2 keys**.

For the decryption part, the procedure follows the same pattern with few more concerns, comparing with the plain DES or Triple DES decryption, regarding the order of the input blocks. This is a very important issue because not using the proper order of the keys would determine the whole decryption process to get into wrong processing.

III TESTING AND OBSERVATIONS

The classic DES is commonly implemented in hardware, especially because the process takes place at the machine level (binary). In the 2K-DES implementation, we converted a hardware application of DES into a software interface for the simulation purpose. The input is being fed from a file. As the software approach requires the initial input is a plain text user friendly file (INP-file).

The input file is converted into a binary data file (we will call it BNE-file - Binary Not Encrypted file) and from now on practically the hardware DES algorithm is simulated in a software representation. Over the BNE-file two keys of 56 bits each are applied, alternatively, for each data block of 64 bits.

The blocks are saved into a cipher file (CIP-file) in the same order in which they were processed, this is very important for the blocks synchronization. The CIP-file represents the encrypted data which can be stored or sent.

We need to recall that DES is a private key block cipher algorithm and as we could see 2K-DES keeps this property, also.

In the same way, for the decryption part, the receiver of the CIP-file applies the same key for each block of data that the block was encrypted with. The resultant file represents the decrypted binary (BDE-file) version of the data and it should be exactly the same as BNE-file, if the decryption was successfully. From now on the conversion from binary to plain text can be easily realized.

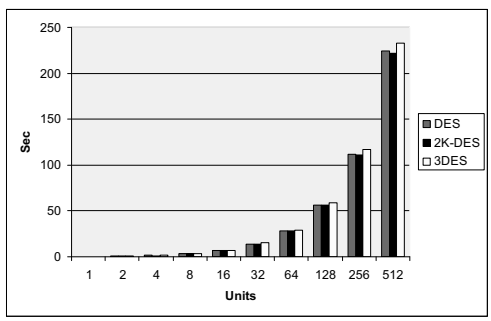
Testing results:

Running the program for each encryption algorithm in turn offers the chance to observe few characteristics regarding of how 2K-DES acts comparing with simple DES and Triple-DES.

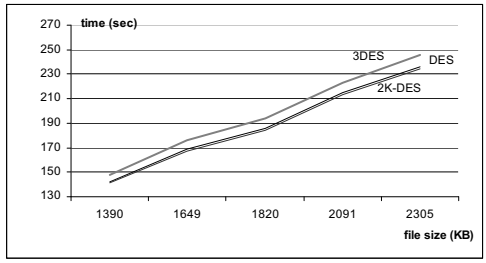
For small input file with binary data all of the tree encryption algorithms are working the same, there were no differences in processing time for the same file. Going further and increasing the quantity of data for encryption, a proportional increase in the Triple-DES encryption processing time is obvious.

On the other hand, for bigger data files, simple DES and 2K-DES are acting similarly.

For these encryption algorithms small files are not relevant in performance analysis because the processing time for encryption is short and is getting lost among the other functions that are implemented. As bigger the file become as accurate the Triple-DES increasing time can be distinguished.

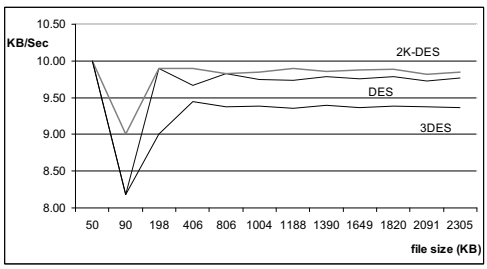


Fig, 2 *DES, 2K-DES and Triple-DES time consuming*



Fig, 3 *DES, 2K-DES and Triple-DES for different size files of data*

Calculating the number of KB encrypted in 1 second (KB/sec) can be noticed that, for big enough files, this report is relatively constant. That means all three encryptions are linear.



Fig, 4 *DES, 2K-DES and Triple-DES – number of KB processed in 1 sec*

$f(x)= ax+b;\quad [f(x)] = sec;\quad [x] = KB;$ (1)

for DES and 2K-DES:

(2)

$$\begin{cases} 215a+b= 2091 \\ 236a+b=2305 \end{cases}$$

Which results in linear function:

$$f(x)=10.2x - 102 \quad (3)$$

for Triple-DES:

(4)

$$\begin{cases} 223a+b= 2091 \\ 246a+b=2305 \end{cases}$$

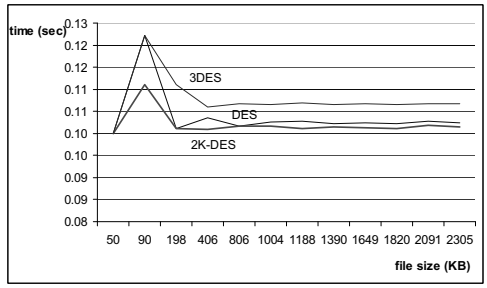
Which results in linear function:

$$f(x)=9.3x +17.2 \quad (5)$$

As we shall see when compute how many seconds are necessary to encrypt 1 KB of data:

	File 4	File 5	File 6	File 7	File 8
KB	406	806	1004	1188	1390
DES (sec)	0.10	0.10	0.10	0.10	0.10
2K-DES (sec)	0.10	0.10	0.10	0.10	0.10
3DES (sec)	0.11	0.11	0.11	0.11	0.11

Fig, 5.1 *DES, 2K-DES and Triple-DES Number of seconds necessary to encrypt 1KB of data*



Fig, 5.2 *DES, 2K-DES and Triple-DES – number of seconds necessary to encrypt 1KB of data*

Unexpected, sometimes 2K-DES is even faster than DES. This might be the result of the software implementation in C++ which might be faster for 2K-DES, even though the same pattern of coding was used for 2K-DES as well as for DES and Triple-DES.

In DES coding in the body of the loop just one block of text is encrypted. In 2K-DES loop 2 blocks data are processed that means, when the end of the loop is reached once, for DES, just one block is done and for 2K–DES two blocks are encrypted. In conclusion the time needed to pass the encryption loop once is shorter for DES, because 2K-DES needs to process two blocks, but overall the total time accumulated is sometimes shorter for 2K-DES.

This, probably, will not be eloquent for a hardware implementation of these algorithms. Still the important remark is Triple-DES is far more time consuming as DES and 2K-DES. 2K-DES inherits DES's high performance and improves the security by doubling the effort required to crack a DES key.

IV. PROS AND CONS 2K-DES

Pros 2K-DES

As a prototype, the first overview on 2K-DES reveals a few interesting characteristics:

Comparing with Triple DES which uses 3 different keys for the same block of plain text, 2K-DES employs only one key for a block of data even though the overall number of keys is two. The main reason of Triple DES's slow performance is represented by the repetitive use of the keys on

the same block which increases significantly the time of the encryption process.

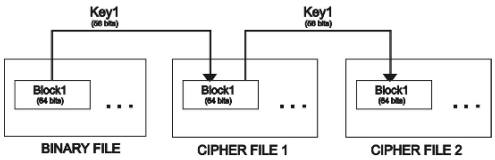


Fig. 6 *Double-DES mode of operation*

Moreover, comparing with a Double-DES algorithm the performance still can be high for 2K-DES because a Double-DES encryption would require for a single text block to be encrypted with the same key twice. As we can logically deduct, a Double-DES means encryption of a 64 bits binary block of data with a 56 bits randomly generated key twice and decryption is reverse: decrypting with the same key twice. In order to avoid this double or triple encryption action(for Double and Triple-DES) on the same data block, 2K-DES might use the simplicity of a single key and keep the higher overall security implied by two keys.

On the other hand, the security weakness of a classic DES could be covered by making the cracker's work more difficult. Probably a 2K-DES crake machine would need twice the brute force needed for cracking DES or twice the time needed for it, because, in fact, it needs to "guess" 2 different keys. Still, the main reason for which the 2K-DES prototype was brought into discussion remains the efficient use of a single traditional DES key for a specific block of data.

Another plus for the new algorithm, but not the least important one is the hardware implementation. As we mentioned before, the complexity and financial side of the algorithm implementation can determine its reliability. A 2K-DES device will be probably built as a normal DES machine which does two DES procedures, each with different generated keys, applied alternatively. The DES encryption machine is cheap and simply enough to be easily built [12]. Practically 2K-DES might be implemented in hardware as a DES device with a little more memory and more care for synchronization. Considering the cost of the new AES (Advanced Encryption Standard) implementation, the 2K-DES might be found more attractive.

Cons 2K-DES

A 2K-DES attempt might not be such a brilliant development because of a few concerns:

The machine that might crake 2K-DES wouldn't be hard to build considering that a DES machine already exists. In the best case it will be probably twice as expensive, or maybe, will just require more time, probably double time as a DES cracking brute-force machine.

As we presented before, for a 2K-DES algorithm the security provided by its key is not doubled compared with DES algorithm. 2K-DES has 2 keys but the length of a 2K-DES key is still 56 bits. We cannot consider the overall length of a 2K-DES being 112 bits, as we are tempted, because the keys are applied independently over different blocks. Triple-DES has indeed a double or triple key length [16] (112 or 168 - Triple ECB, Triple CBC [17]) because two or respectively three keys are applied to the same input block and their security cumulate.

Variations on the Same Theme

Obviously, like 2K-DES, a 3K-DES or why not a 4K-DES could be developed also. The question that would arise in that case might be: The performance of these methods is better than the Triple DES? Theoretically the answer is yes at least for 3K-DES because each block is encrypted, respectively decrypted with a single key, the only extra time spent being for generating one more key.

V. CONCLUSIONS:

Summary:

Summarizing the work-flow that has been made to obtain a performance comparison of these three encryption algorithms in software, we can say that a similar program for DES, 2K-DES and Triple-DES, running for the same file shows a lower performance for Triple-DES. Working with files in software implementation shows an evident more efficient 2K-DES algorithm comparing with Triple-DES.

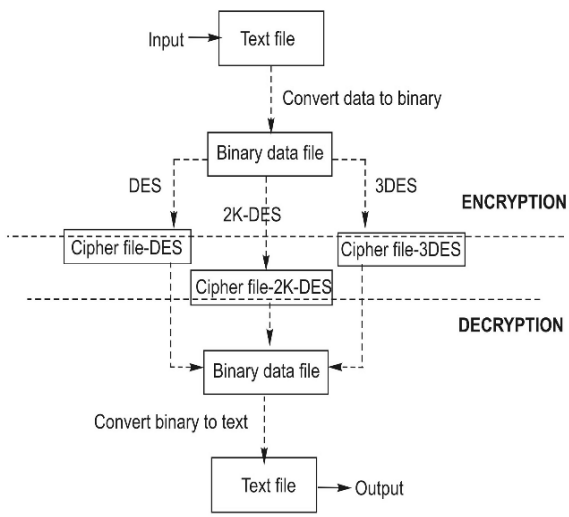


Fig. 7 *C++ Implementation flowchart*

During the encryption / decryption process only the cipher file has changed. The other files, binary or text, are the same for all algorithms and require the same amount of time to be generated.

Then we had only one variable file – Cipher file and its creation time was measured for DES, 2K-DES and Triple-DES.

Security versus performance:

From the security point of view 2K-DES is not such a huge improvement compared to Triple-DES, because in fact, security is not even doubled comparing with simple DES algorithm. Still, if we take in consideration that Triple–DES does not significantly increase security over double encryption [4] the security performance can be compared with a double encryption because for either one, 2 keys are generated and, implicitly, in a cracking process, they need to be found. If the probability to find a DES key is 2^{64}= 72,000,000,000,000,000 [11], than the work required by 2 keys is 2×2^{64} . But this is true just for Double-DES encryption. In this case the keys' effects cumulate because they are applied over the same block of data. It is not the case for 2K-DES, as we might be tempted to believe but the security of 2K-DES is considerably higher comparing with simple DES. It can be still considered an improvement of simple DES in terms of security.

Future work:

The implementation of 2K-DES is just an idea and a detailed research hasn't been made so far. It would be an interesting issue to develop a hardware implementation for this algorithm. As we know DES implementation is much faster in hardware than in software [8], the performance of future implementation of 2K-DES in hardware will be significantly better because practically 2K-DES is a simple DES used twice.

On the other hand, future software research could have the topic: "Double encryption with 2K-DES". This algorithm could combine Double-DES and 2K-DES as follows:

The first input block is encrypted with the first random generated key. The resultant is encrypted again with a second key.

The next block is encrypted first with Key2 and then with Key1. Using just 2 keys the effort to find these keys is very much higher than Double-DES or 2K-DES. This could be the main advantage of this kind of implementation. Other combinations can be found in the search of the one with the best security and performance.

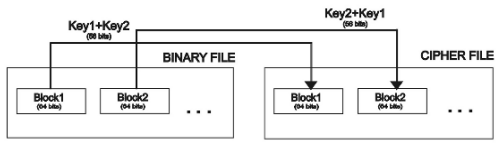


Fig. 8 *Double-DES with 2K-DES*

Regardless of how many years DES will survive as a safe enough encryption algorithm the main idea is that, anytime, something different could be found, tested and why not, sometimes used.

REFERENCES:

[1] AvnetCore : Datasheet
Data Encryption Standard – Version 1.0 July 2006
[2] D Coppersmith The Data encryption Standard (DES) and its Strength against attacks – IBM 1 RES DEVELOP vol 38 no 3 May 1994
[3] Burt Kalinsky – Article "A survey of Encryption Standards " IEEE Micro December 1993
[4] W. Diffie and M. E. Hellman, "Exhaustive Cryptanalysis of the NBS Data Encryption Standard," in IEEE Computer, vol. 10, 1977, pp. 74-84. (not on the web)
[5] M. Wiener, "Efficient DES key search", TR-244 Carleton University, 1993 (also appears in William Stallings, Practical Cryptography for Data Internetworks, IEEE Computer Society Press, 1996)
(also appeared in the proceedings of CRYPTO 1993)
[6] Wing Ng Wen / Chun Yang
Use DES to generate random numbers in C programming language-Oregon State University
[7] Laissez Faire City Times, Vol 2, No. 28.
Homepage: http://orlingrabbe.com/
Laissez Faire City Times: http://zolatimes.com/
[8] S.Bitan, , D.Frommer, RADGUARD
PSEC Working Group Internet Draft August 1997
[9] http://www.nic.funet.fi/pub/crypt/mirrors/ftp.dsi.unimi.it/docs/des-how-to.txt
[10] http://www.quadibloc.com/crypto/co040201.htm
http://ecommerce.hostip.info/pages/288/Data-Encryption-Standard-DES.html
[11] http://www.laynetworks.com/users/webs/des.htm
[12] http://www.disappearing-inc.com/D/des.html
[13] http://www.interhack.net/projects/deschall/
[15] Tom Kean, Ann Duncan Xilinix Development Corporation, Scotland " DES key breaking, Encryption and Decryption on the XC62 16
[16] http://searchsecurity.techtarget.com/tip/1,289483,sid14_gci968714,00.html?track=NL-102&ad=486202
[17] Department of Commerce-National Institute of Standards and Technology "Announcing Draft Federal Information Processing Standard (FIPS) 46-3,Data Encryption Standard (DES), and Request for Comments" - Federal Register: January 15, 1999

One-Time Password Authentication with Infinite Hash Chains

Alexander G. Chefranov
Eastern Mediterranean University, Gazimagusa, North Cyprus
Institute of Technology, South Federal University, Taganrog, Russia

***Abstract*- One-time password scheme based on infinite forward stepping hash chains and not requiring re-initialization after a certain number of authentications is proposed. It resists known for such type schemes attacks, channel failures, and server side database compromise.**

I. Introduction

Authentication is an important problem that can be solved by one time password (OTP) schemes such as [1]-[3] which assume not secure channels of communication between clients and servers, and possible compromise of passwords on the server side. Also, approach [3] is robust to possible channel failure. These OTP schemes use hash chains

$$h(h(..h(p)..)) = h^n(p) \quad (1)$$

applied to a secret password phrase *p*, where *n* is the length of the hash chain. The hash function *h* is assumed to be a one-way function [4] that is easy to calculate on the given argument, and is computationally infeasible to invert. Hash chains have a property of easy stepping forward (to step from $h^n(p)$ to $h^{n+1}(p)$), but it is not possible to go backward without knowledge of the secret information. OTP schemes [1]-[3] based on backward chains have two major problems: 1) significant computational complexity imposed by the protocols on a client side due to necessity of hash function exponentiation according to (1), and 2) necessity of re-initialization of a password phrase after some number of authentications. Also, such schemes are vulnerable to the small number attack, in which a fake server asks a client for a challenge $h^s(p)$, where *s* is a small number. Having such a value, an opponent can easily calculate response on an actual server challenge requiring greater hash function powers. Reducing of the client side computational effort is achieved in [1] by the use of restricted hash chains allowing *R* forward steps without knowledge of the secret information (RHC-*R*); re-initialization after a certain number of authentication is still required. An OTP scheme using asymmetric encryption for hash function calculation and providing not limited number of uses before re-initialization is suggested in [5]. It is subject to compromise of the client's public key which is to be kept on the server side; the server is not authenticated to the client. OTP scheme [6] uses mutual authentication of a client and server; it uses simple XOR-encryption and is resistant to the small number attack but it requires re-initialization after a limited number of authentications and uses long backward hash chains. Scheme [6] is vulnerable to preplay attack [7]. We propose a one-time password scheme based on RHC-1; the scheme does not need re-initialization after a certain number of authentications, it has better performance than that of other OTP schemes, and it is not vulnerable to known for OTP attacks. The rest of the paper is organized as follows. Section 2 presents details on some OTP schemas. In Section 3, we introduce our OTP scheme using infinite RHC-1. Conclusions are presented in Section 4.

II. Some OTP Schemas

A. Lamport's scheme

Backward hash chains [3] are used as follows. In initialization step, a user securely supplies a server with a server's password $P_N(p) = h^N(p)$. For *t*-th authentication, the server sends a challenge to the user as a number

$$C_t = N - t, \quad (2)$$

to which the user responds by calculation of an authenticator

$$A_{N-t}(p) = h^{N-t}(p), \quad (3)$$

$t > 0$ - integer. For *t=1*, the user responds with the authenticator $A_{N-1}(p) = h^{N-1}(t)$, and server's password currently is $P_N(p) = h^N(p)$. The server authenticates the user by checking that the following equality holds:

$$h(A_{N-t}(p)) = P_{N-t+1}(p). \quad (4)$$

After successful authentication, the server sets

$$P_{N-t}(p) = A_{N-t}(p), \quad (5)$$

and for the next authentication, the server will increment *t* and send the new challenge C_t (2). Thus, we see that for OTP [3], the authentication process starts with the large power, *N*, and proceeds by successive power decrementing. Authenticators,

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 283–286.

$A_i(p)$, are calculated by hash function exponentiation (3), and after successful authentication according to (4), they are used as next server side passwords, P_i, $i = \overrightarrow{N-1,1}$. After *N* authentications, re-initialization is required. This schema is vulnerable to the small number attack, because, if an opponent would send to the client a small number, say, *s=10*, as a server's challenge, the client will respond by an authenticator $A_{10}(p) = h^{10}(p)$, using which the opponent can easily calculate all authenticators with greater numbers, and will be able to impersonate the user in communications with the true server, providing proper responses on the server's challenges. Another problem of OTP [3] is significant computational effort of the client for calculation of hash function powers (3) (number of hash function calculations for *t*-th authentication is *N-t+1*).

B. Scheme of Goyal et. al

A way of decreasing a client's computational load is proposed in [1] where the large authentication period *N* is split into *N/R* sub-periods of length *R* (let's consider for simplicity that *N* is a multiple of *R*) . The hash chain is constructed in such a way, that stepping forward inside a sub-period is possible without knowledge of the secret information, but crossing a boundary between sub-periods requires such information. This is achieved by the use of the following formula:

$$A_{k+1}(p) = h(A_k(p) + \delta \cdot p), \quad (6)$$

where $\delta = \begin{cases} 1, & k \bmod R = 0 \\ 0, & else \end{cases}$, $A_0(p) = p$, *p* is the secret password, $k \geq 0$. Authentication proceeds as in the case of backward hash chains [3] but with the following distinctions allowing to decrease computational effort of the client and to counter the small number attack in some extent. Values of A_{iR} are supplied to the server in initialization step, and they serve as server passwords, P'_i, $i = \overline{1, N/R}$; the current server password is $P_N = A_N = P'_{N/R}$. For *t*-th authentication, the server computes $n = (N-t) \bmod R$ and $k = ((N-t) divR)R$. Obviously, *n* shows displacement of *t*-th point in the k/R -th sub-period:

$$N - t = \frac{k}{R} \cdot R + n. \quad (7)$$

Challenge sent by the server to the user has a form

$$C_{n,k} = (n, A_k(p)). \quad (8)$$

User responds by calculation of the authenticator, based on (6) and (7):

$$A_{N-t}(p) = h^n(A_k(p) + p). \quad (9)$$

Authentication, as for hash chains [3], is made by the use of (4), and the new current server password is set by use of (5). Distinction between scheme [3] and this schema is in the following: the latter scheme restricts stepping forward without knowledge of the secret password, *p*, to *R* steps; hence, if *R* is not large, number of fake authentications due to the small number attack is limited to *R*. Hence, it is desirable to have *R* as minimal as possible. But, from the other side, decrease of *R* means increase of the number server passwords, and in the limit, all *N* passwords are to be kept in the server's database; this means vulnerability due to possible server's database compromise. Backward hash chain schema [3] is resistant to the server's database compromise. That's why *R* is limited from below. Not large values of *R* allow restricting client's computational effort from above due to (7) and (9) but *n* may be also large, causing significant client's computational effort (number of hash function calculations per authentication on average is *n/2*). Re-initialization after *N* authentications is still necessary as for the scheme [3].

C. Bicakci and Baykal's scheme

Infinite length hash chains [5] are free of necessity of re-initialization after a certain number of authentications. This chain is constructed by consecutive public key encryption, and checking is performed by respective decryption which should output some secret known value. The algorithm is time consuming due to used in it exponentiations.

D. Yeh-Shen-Hwang's scheme

We follow [7] in description of [6]. In the protocol below we denote Client as C, and Server as S. The protocol has two stages: Registration, and Login and authentication. We describe stages of the protocol as a sequence of actions. Each action is performed by S or C that is specified at the beginning of an item; $C \leftarrow S : M$ means that message *M* is transmitted from Server to Client; transmission in the opposite direction is shown similarly.

1. Registration stage

1.1. $C \leftarrow S$: *SEED*

1.2. $C \leftarrow S : N,\ SEED \oplus D_0,\ h(D_0)$

1.3. $C \rightarrow S :\ p_0 \oplus D_0$

where $p_t = h^{N-t}(K \oplus SEED), t \geq 0$, *K* is a Client's secret password phrase, $\oplus$ is a bitwise XOR operation, *N* is the number of login times before re-initialization, $N >> 1$, $SEED$ is a random number generated by Server and securely delivered to Client, D_0 is the first nonce (random number) generated by Server. Messages

of steps 1.2 and 1.3 can be sent via not secure channels.

2. Login and authentication stage

2.1. $C \leftarrow S$: M1=(M_{11}, M_{12}, M_{13}) =($N-t, SEED \oplus D_t, h(D_t) \oplus p_{t-1}$), $N-1 \geq t \geq 1$ - integer.

2.2. C: if $h(M_{12} \oplus SEED) \oplus p_{N-M_{11}-1} \neq M_{13}$ then Client decides that message M1 is not from Server and stops the procedure.

2.3. $C \rightarrow S$: $U_t = p_t \oplus D_t$

where p_t is a calculated by Client new one-time password that should replace the previous one, p_{t-1} on the server side.

2.4. S: if $h(U_t \oplus D_t) = p_{t-1}$ then Server authenticates Client, where p_{t-1} is one-time password that is already kept by Server. Finally, the new password, $p_t = U_t \oplus D_t$, replaces p_{t-1} on the server side, and *t* is incremented.

This scheme is vulnerable to preplay attack [7] because password information goes in two ways: from the server to the client, and from the client to the server. An attacker potentially has possibility to impersonate the server to the client (by sending the client a forged challenge), and after that to impersonate the client to the server using a new valid password-authenticator sent to him by the user, that was demonstrated in [7]. We see also that scheme [6] is practically the same as [3] but uses closing of sensitive parts of transmitted messages with the help of XOR operation, hash function, and random nonces. Number of hash function calculations for *t*-th authentication is *N-t+4*.

III. One-Time Password with Infinite RHC-1

Our OTP scheme uses RHC-1 to build infinite forward hash chains that excludes necessity of 1) re-initialization, and 2) client-side exponentiations (3). We use hash functions, XOR operation, and nonces to hide sensitive parts of messages and to provide authenticity of messages similarly to how it is made in [6]. In the protocol below, a client conveys to a server not only its authenticator, but also a server-side password for the next authentication session. The proposed OTP scheme generates RHC-1 in the forward direction contrary to schemes such as [1], [3], [6] that pass hash chain in backward direction (from the last element to the first one).

The Protocol.

1. Registration stage

1.1. S: $n_S = 0$; generate randomly $SEED_S$

1.2. $C \leftarrow S$: $SEED_S$ (secure transmission)

1.3. C: $n_C = 0$; generate randomly $SEED_C$

1.4. $C \rightarrow S$: $SEED_C$ (secure transmission)

1.5. C: $p = h(K \oplus SEED_S)$, $p_1 = h(p)$, $hp = h(p_1)$

1.6. $C \rightarrow S$: *hp* (secure transmission - S gets the first server side password *hp*)

2. Login and authentication stage

2.1. C: generate nonce D_C

2.2. $C \rightarrow S$: M1=($M_{11}, M_{12}, M_{13}, M_{14}$) = ($p_1 \oplus h(D_C)$, $h(p \oplus h(p_1)) \oplus h(D_C)$, $h^2(p \oplus h(p_1)) \oplus D_C$, $SEED_C \oplus D_C$)

2.3. S: $D_C = SEED_C \oplus M_{14}$;

If $h(M_{11} \oplus h(D_C)) = hp$ and $h(M_{12} \oplus h(D_C)) = M_{13} \oplus D_C$, then Server authenticates Client, else – Server stops the protocol

2.4. S: $V = D_C + 1$, $hp = M_{12} \oplus h(D_C)$, $n_S = n_S + 1$

2.5. $C \leftarrow S$: M2=($V \oplus SEED_S, V \oplus SEED_C$) = (M_{21}, M_{22})

2.6. C: if $M_{21} \oplus SEED_S = M_{22} \oplus SEED_C = D_C + 1$ then Client decides that authentication is completed successfully, otherwise the protocol stops

2.7. C: $p_1 = p \oplus h(p_1)$, $n_C = n_C + 1$.

In the protocol, Client randomly generates nonce D_C, and hash of a new password of the client $h(p \oplus h(p_1))$ = *h(new password)*, and sends message M1 to Server. Server, upon receiving M1, reveals $D_C = SEED_C \oplus M_{14}$, $p_1 = M_{11} \oplus h(D_C)$, and *h(new password)*$= M_{12} \oplus h(D_C)$, calculates $h(p_1)$, and compares it against *hp* kept in its database. Authenticity of M1 is checked by comparison of $h(M_{12} \oplus h(D_C))$ against $M_{13} \oplus D_C$. If the conditions hold, Server decides that a client is actually Client. In this case, Server sets $V = D_C + 1$, *hp=h(new password)*, $n_S = n_S + 1$, and sends the message M2 to Client. Client, on receiving M2, reveals *V* and compares it against $D_C + 1$. Authenticity of M2 is checked by: $M_{21} \oplus SEED_S = M_{22} \oplus SEED_C = D_C + 1$. If the conditions hold, Client sets $n_C = n_C + 1$, $p_1 = p \oplus h(p_1)$, and authentication completes successfully.

Considered above scheme is not limited to a finite

number of interactions before a password re-initialization, has the low computational complexity both on a client and server sides. Performance of the suggested protocol is defined by four hash function calculations on the client side (Step 2.2), and three hash function calculations on the server side (Step 2.3). Thus, the total number of hash function calculations per authentication is seven. Scheme [6], for *t*-th iteration, requires *N-t+4* hash function calculations. Similar estimates hold for other OTP schemas such as [3]; better performance shows [1] but still it is worse than that of the presented here schema.

The protocol is resistant to compromising of a server side password database because the server keeps only hash of the current client's password. New passwords are generated by a client's side only with the simple procedure (client side password in Step 2.7, and server side password in Step 2.2).

The procedure of password generation may be represented as follows:

1. Initial settings: $p = h(K \oplus SEED_S)$. Client side password: $p_1 = h(p)$; server side password: $hp_1 = h(p_1)$

2. Next passwords generation. Client side password: $p_{t+1} = p \oplus h(p_t)$; server side password: $hp_{t+1} = h(p_{t+1})$, $t = 1,2,..$

Replay attack on the scheme is not possible since each time password changes. Small number attack is not possible because a challenge (an authentication number) is not sent by a server, and forward hash chains are used. Preplay attack is not possible since password information goes only in one way, from a client to a server. A server is authenticated by a client with the help of nonce D_C: only the server could understand value of D_C, increase and close it with known to the server $SEED_C$ and $SEED_S$.

The scheme is resistant also to channel failures. If Client does not get a response M2 from the server, it might happen due to failure of delivery of the message M1 to the server, or because of failure of delivery of M2 to the client from the server. In such a case, the client can resolve the uncertainty by sending to the server request for $V' = n_s$. If $n_c + 1 = V'$, it means that the server has received M1. In such a case, Client can complete authentication successfully by performing Step 2.7. Otherwise, M1 is lost, and its copy should be sent again.

IV. Conclusion

Thus far, we presented a new one time password protocol based on RHC-1 and having significantly better performance than that of known OTP schemas. The proposed schema is not limited to a certain number of authentications before re-initializations contrary to known OTP schemas based on hash chains, and does not use complicated cryptographic algorithms contrary to OTP schemes, such as [5], providing infinite number of authentications. Total number of hash function calculations per authentication is seven. Messages between sides are transferred in the closed form using XOR operation; message authenticity is supported by use of hash function digests. The protocol is resistant to the small number attack because a challenge (an authentication number) is not sent by a server, and forward hash chains are used. It is resistant to the preplay attack that is achieved by one-way transfer of password information. To counter channel failures, the restoration procedure is introduced. Proposed in the scheme the procedure of a password generating from a single secret password phrase may be used in applications where a client communicates with multiple servers each of them requiring its own password; the number of such servers and time of the use are not limited.

References

[1] V. Goyal, A. Abraham, S. Sanyal, and S.Y. Han, "The N/R one time password system." in *Proceedings of International Conference on Information Technology: Coding and Computing (ITCC'05)*, 4-6 April, 2005, vol. 1, pp. 733-738, 2005.

[2] N. Haller, "The S/KEY one-time password system." *RFC 1760*, 1995, available from http://www.ietf.org.

[3] L. Lamport, "Password authentication with insecure communication." *Comm. ACM,* vol. 24, No 11, pp. 770-772, 1981.

[4] W. Diffie and M.E. Hellman, "New directions in cryptography." *IEEE Trans. Inform. Theory,* vol. IT-22, No 6, pp. 644-654, 1976.

[5] K. Bicakci and N. Baykal, "Infinite length hash chains and their applications." in *Proceedings of 11th IEEE Int. Workshops on Enabling Technologies: Infrastructure for Collaborating Enterpri*ses (WETICE'02), pp. 57-61, 2002.

[6] T.C. Yeh, H.Y. Shen, and J.J. Hwang, "A secure one-time password authentication scheme using smart cards." *IEICE Trans. Commun.,* vol. E85-B, No. 11, pp. 2515-2518, 2002.

[7] D.H. Yum and P.J. Lee, "Cryptanalysis of Yeh-Shen-Hwang's one-time password authentication scheme." *IEICE Trans. Commun.,* vol. E88-B, No. 4, pp. 1647-1648, 2005.

Estimation of OFDM Time-Varying Fading Channels Based on Two-Cross-Coupled Kalman Filters

Ali Jamoos, Ahmad Abdo, and Hanna Abdel Nour

Department of Electronics Engineering, Al-Quds University, P.O. Box 20002, Jerusalem, Palestine
Email: ali@eng.alquds.edu; aabdo@eng.alquds.edu; habdalnour@eng.alquds.edu

***Abstract*—This paper deals with the estimation of rapidly time-varying Rayleigh fading channels in Orthogonal Frequency Division Multiplexing (OFDM) mobile wireless systems. When the fading channel is approximated by an Autoregressive (AR) process, it can be estimated by means of Kalman filtering. Nevertheless, the AR model order has to be selected. In addition, the AR parameters must be estimated. One standard solution to obtain the AR parameters consists in first fitting the AR process autocorrelation function to the theoretical Jakes one and then solving the resulting Yule-Walker Equations (YWE). However, this approach requires the Doppler frequency which is usually unknown. To avoid the estimation of the Doppler frequency, the joint estimation of both the channel and its AR parameters can be addressed. Instead of using the Expectation-Maximization (EM) algorithm which results in large storage requirements and high computational cost, we propose to consider a structure based on two-cross-coupled Kalman filters. It should be noted that the Kalman filters are all the more interactive as the variance of the innovation of the first filter is used to drive the Kalman gain of the second. Simulation results show the effectiveness of this approach especially in high Doppler rate environments.**

***Index Terms*—Autoregressive processes, Rayleigh fading channels, Jakes model, Kalman filters, OFDM.**

I. INTRODUCTION

Orthogonal Frequency Division Multiplexing (OFDM) is a parallel data transmission scheme in which high data rates can be achieved by the simultaneous transmission over many orthogonal carriers [1][2]. This multi-carrier transmission scheme makes it possible to convert the severe wide-band frequency-selective fading channel into many narrow-band frequency non-selective flat fading sub-channels, which are free from Inter-Symbol Interference (ISI). Due to the various advantages of OFDM, it has been adopted in many wide-band digital communication systems such as digital audio and video broadcasting, Asynchronous Digital Subscriber Lines (ADSL), IEEE 802.11 a/g Wireless Local Area Networks (WLAN), etc. [3].

In OFDM systems, due to user mobility, each carrier is subject to Doppler shifts resulting in time-varying fading. Thus, the estimation of the fading process over each carrier is essential to achieve coherent symbol detection at the receiver. In that case, training sequence/pilot aided techniques and blind techniques are two basic families for channel estimation [4]. In this paper, as blind techniques require longer observation window and have higher complexity than training based techniques, we will focus our attention on training based channel estimation techniques.

The time-varying Rayleigh fading channels are usually modeled as zero-mean wide-sense stationary circular complex Gaussian processes, whose stochastic properties depend on the maximum Doppler frequency denoted by f_d. According to the Jakes model [5], the theoretical Power Spectrum Density (PSD) associated with either the in-phase or quadrature portion of the fading process, is band-limited and U-shaped. Moreover, it exhibits twin peaks at $\pm f_d$. This key information about channel statistics cannot be however exploited when directly estimating the fading processes by means of the Least Mean Square (LMS) and the Recursive Least Square (RLS) algorithms as in [6]. Alternatively, Kalman filtering combined with an Autoregressive (AR) model to describe the time evolution of the fading processes is shown to provide superior Bit Error Rate (BER) performance over the LMS and RLS based channel estimators [6] [7]. Nevertheless, the AR model order has to be selected. In addition, the AR model parameters are unknown and, hence, must be estimated.

On the one hand, several authors (e.g., [7][8][9][10]) have expressed the AR parameters by first fitting a low-order AR process autocorrelation function to the theoretical Jakes one and then solving the resulting Yule-Walker Equations (YWE). However, this requires the preliminary estimation of the maximum Doppler frequency, which is not necessarily a trivial task [11].

On the other hand, the AR parameters can be estimated from the received noisy signal. Among the existing methods, Tsatsanis *et al.* [12] have proposed to estimate the AR parameters from estimates of the channel covariance function, by means of a standard Yule-Walker estimator. However, this method results in biased AR parameter estimates. In [13], Cai *et al.* have proposed a channel estimation scheme for OFDM wireless systems based on two-serially-connected Kalman or H_∞ filters. The first one is used for AR parameter estimation and the second one for fading process estimation. Nevertheless, the AR parameter estimates are biased since they are estimated directly from the noisy data. This might result in poor estimation of the fading process. To avoid this drawback, one can look at other approaches proposed in other fields than mobile communications. Thus, the Expectation-Maximization (EM) algorithm which often implies a Kalman smoothing could be used [14]. Nevertheless, since it operates repeatedly on a batch of data, it results in large storage requirements and high computational cost. In addition, its success depends on the initial conditions. As an alternative, two recursive filters could be cross-coupled to solve the so-called dual estimation issue [15], i.e. the estimations of both the AR process and its parameters. Each time a new observation is available, the first filter uses the latest estimated AR parameters to estimate the signal, while the second filter

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 287–292.

uses the estimated signal to update the AR parameters. According to Gannot [16], this dual filtering approach can be viewed as a sequential version of the EM algorithm. Recently, a variant [17] based on two-cross-coupled Kalman filters has been developed in which the variance of the innovation process in the first filter is used to define the gain of the second filter. As this solution can be seen as a recursive instrumental variable technique, consistent estimates of the AR parameters are obtained. Meanwhile, we have analyzed the relevance of this approach to estimate Multi-Carrier Direct-Sequence Code Division Multiple Access (MC-DS-CDMA) fading channels in [18].

In this paper, we propose to take advantage of the two-cross-coupled Kalman filters for the joint estimation of time-varying OFDM fading channels and their corresponding AR parameters. See Fig. 1. In addition, both low and high order AR models are investigated.

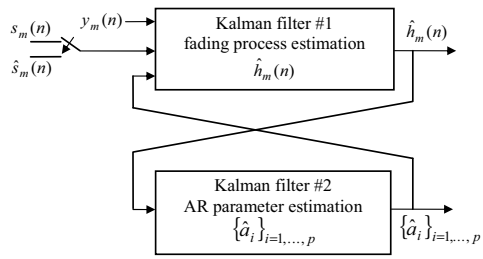


Fig. 1. Two-cross-coupled Kalman filter based structure for the joint estimation of the fading process and its AR parameters along the m^{th} carrier.

The remainder of the paper is organized as follows. In section II, we recall the OFDM system model. The estimation of the fading channels based on two-cross-coupled Kalman filters with high-order AR models is introduced in section III. Simulation results are reported in section IV. Conclusion remarks are drawn in section V.

II. OFDM System Model

Let us consider a standard OFDM system as shown in Fig. 2. In this system, the input serial data stream is firstly converted into parallel data blocks. An Inverse Fast Fourier Transform (IFFT) is performed on each block and a guard interval to eliminate the Inter-Symbol Interference (ISI) is then added before they are transmitted through the channel. The transmitted OFDM signal is assumed to go through a rapidly time-varying frequency-selective Rayleigh fading channel. At the receiver, the FFT is performed on each received OFDM symbol after the guard interval being removed. Thus, with proper selection of the guard interval and perfect carrier synchronization, the received signal sample over the m^{th} carrier for the n^{th} OFDM symbol can be written in the following manner

$$y_m(n) = h_m(n)s_m(n) + w_m(n), \qquad m = 1, 2, \ldots, M \tag{1}$$

where $s_m(n)$ is the m^{th} information symbol of the n^{th} OFDM symbol, $h_m(n)$ is the fading process over the m^{th} carrier during the n^{th} OFDM symbol, $w_m(n)$ is an Additive White Gaussian Noise (AWGN) process, and N is the total number of carriers.

The information symbols $s_m(n)$ are assumed to be drawn from Quadrature Phase shift Keying (QPSK) constellation $\{1, -1, j, -j\}$ independently for different m and n. The noise processes $\{w_m(n)\}_{m=1,\ldots,M}$ are assumed to be mutually independent and identically distributed zero-mean complex Gaussian processes, with equal variance σ_w^2. The fading process over the m^{th} carrier $h_m(n) = \beta_m(n)e^{j\theta_m(n)}$ is assumed to be a zero-mean complex Gaussian process with uniformly distributed phase $\theta_m(n)$ on $[0, 2\pi[$ and a Rayleigh distributed envelop $\beta_m(n)$. The variances of the fading processes $\{h_m(n)\}_{m=1,\ldots,M}$ are all assumed equal to σ_h^2.

The stochastic characteristics of the m^{th} carrier fading process $h_m(n)$ depend on the maximum Doppler frequency:

$$f_d = v f_c / c \tag{2}$$

where v is the mobile speed, f_c is the central carrier frequency and c is the light speed.

According to [5], the theoretical Power Spectral Density (PSD) associated with either the in-phase or quadrature portion of the fading process $h_m(n)$ is band-limited and U-shaped. Moreover, it exhibits two peaks at $\pm f_d$ as follows:

$$\Psi_{hh}(f) = \begin{cases} \dfrac{1}{\pi f_d \sqrt{1-(f/f_d)^2}}, & |f| \le f_d \\ 0, & \text{else where} \end{cases} \tag{3}$$

Its corresponding normalized discrete-time Autocorrelation Function (ACF) hence satisfies:

$$R_{hh}(n) = J_0\left(2\pi f_d T_s |n|\right) \tag{4}$$

where $J_0(.)$ is the zero-order Bessel function of the first kind, T_s is the symbol period, and $f_d T_s$ denotes the Doppler rate.

III. Kalman Filtering based Channel Estimation

A. AR Modeling of Rayleigh Fading Channels

To exploit the statistical properties of the fading channel given by its PSD and ACF, the fading process over the m^{th} carrier can be modeled by a p^{th} order AR process, denoted by AR(p) and defined as follows [7][19]:

$$h_m(n) = -\sum_{i=1}^{p} a_i h_m(n-i) + u_m(n) \tag{5}$$

where $\{a_i\}_{i=1,\ldots,p}$ are the AR model parameters and $u_m(n)$ denotes the zero-mean complex white Gaussian driving process with equal variance σ_u^2 over all carriers.

However, selecting a low-order AR model for the channel is debatable. Some authors [7][12] choose first or second order AR process, as it is simple and the corresponding computational cost is low. Nevertheless, a deterministic model should be used for the channel due to the band-limited nature of its PSD according to Kolmogoroff-Szego formula [20]:

$$\sigma_u^2 = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\left(\Psi_{hh}(\omega)\right) d\omega\right) \tag{6}$$

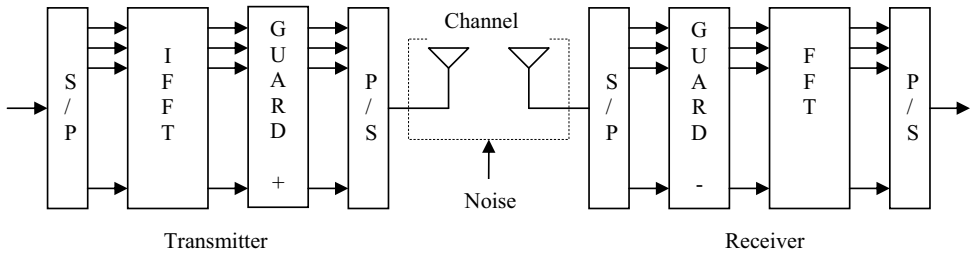


Fig.2: OFDM System.

where $\Psi_{hh}(\omega)$ denotes the PSD of the AR process that fits the theoretical Jakes spectrum. Meanwhile, alternative solutions have been also studied. Firstly, a sub-sampled AR Moving Average (ARMA) process followed by a multistage interpolator has been used by Schafhuber *et al.* [21] in the framework of channel simulation. Nevertheless, only a very high down-sampling factor leads to a PSD which is never equal to 0. Secondly, Baddour *et al.* [19] use high-order AR processes (e.g. $p \geq 50$) when they simulate the channel. For this purpose, they modify the properties of the channel by considering the sum of the theoretical fading process and a zero-mean white process whose variance ε very small (e.g., $\varepsilon = 10^{-7}$ for $f_d T_s = 0.01$). Then, the AR parameters are estimated with the Yule-Walker (YW) equations based on the modified autocorrelation function

$$R_{hh}^{\text{mod}}(n) = J_0\left(2\pi f_d T_s |n|\right) + \varepsilon\delta(n) \tag{7}$$

Taking into account the above discussion, we propose in this paper to use an AR model whose order is high enough to approximate the channel. Fig. 3 and Fig. 4 show, respectively, the autocorrelation function and the power spectrum density of the Jakes model and that of the fitted AR process whose order is 1, 2, 5 and 20.

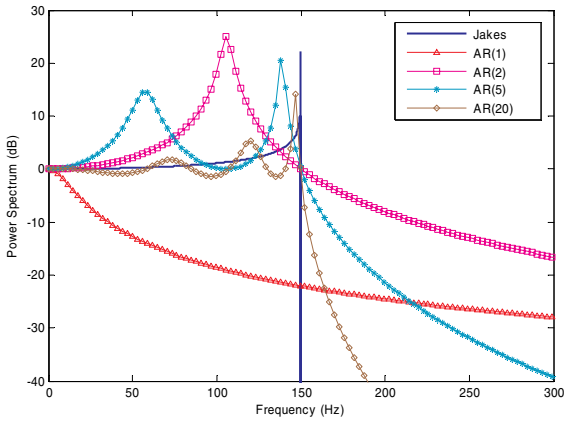


Fig. 3. Power spectral density of the Jakes model and that of the fitted AR (p) process with p=1, 2, 5, and 20. $f_d = 150\,\text{Hz}$ and $f_d T_s = 0.05$.

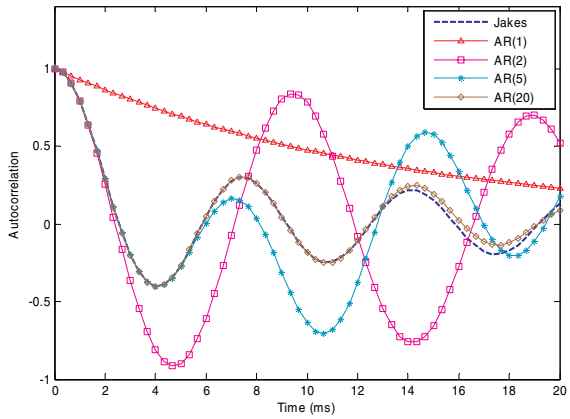


Fig. 4. Autocorrelation function of the Jakes model and that of the fitted AR (p) process with p=1, 2, 5, and 20. $f_d = 150\,\text{Hz}$ and $f_d T_s = 0.05$.

In the following, as $R_{hh}(n)$ is usually unknown, we propose to complete the joint estimation of the fading process $h_m(n)$ and its AR parameters.

A. Estimation of the Fading Processes

To estimate the fading process $h_m(n)$ along the m^{th} carrier, let us define the state vector as follows:

$$\mathbf{h}(n) = \left[h(n) \quad h(n-1) \quad \Lambda \quad h(n-p+1)\right]^T \tag{8}$$

Note that, for the sake of simplicity and clarity of presentation, the carrier subscript is dropped. Then, equation (5) can be written in the following state space form:

$$\mathbf{h}(n) = \mathbf{\Phi}\,\mathbf{h}(n-1) + \mathbf{g}\,u(n) \tag{9}$$

where:

$$\mathbf{\Phi} = \begin{bmatrix} -a_1 & -a_2 & \ldots & -a_p \\ 1 & 0 & \ldots & 0 \\ & \ddots & & \vdots \\ 0 & \ldots & 1 & 0 \end{bmatrix} \text{ and } \mathbf{g} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T \tag{10}$$

In addition, given (1) and (8), one has:

$$y(n) = \mathbf{s}^T(n)\mathbf{h}(n) + w(n) \tag{11}$$

where $\mathbf{s}(n) = \left[s(n) \quad 0 \quad \cdots \quad 0\right]^T$.

Hence, equations (9) and (11) define the state space representation dedicated to the one-carrier fading channel system (1) and (5). At that stage, a standard Kalman filtering algorithm can be carried out to provide the estimation $\hat{\mathbf{h}}(n/n)$ of the state vector $\mathbf{h}(n)$ given the set of observations $\{y(i)\}_{i=1,\dots,n}$ as listed below:

The so-called innovation process $\alpha(n)$ is first obtained:

$$\alpha(n) = y(n) - \mathbf{s}^T(n)\boldsymbol{\Phi}\hat{\mathbf{h}}(n-1/n-1) \tag{12}$$

Its variance is then defined:

$$C(n) = E[\alpha(n)\alpha^*(n)] = \mathbf{s}^T(n)\mathbf{P}(n/n-1)\mathbf{s}(n) + \sigma_w^2 \tag{13}$$

where the so-called *a priori* error covariance matrix $\mathbf{P}(n/n-1)$ can be recursively obtained as follows:

$$\mathbf{P}(n/n-1) = \boldsymbol{\Phi}\mathbf{P}(n-1/n-1)\boldsymbol{\Phi}^H + \mathbf{g}\sigma_u^2\mathbf{g}^T \tag{14}$$

The Kalman gain is calculated in the following manner:

$$\mathbf{K}(n) = \mathbf{P}(n/n-1)\mathbf{b}(n)C^{-1}(n) \tag{15}$$

The *a posteriori* estimate of the state vector and the fading process are respectively given by:

$$\hat{\mathbf{h}}(n/n) = \boldsymbol{\Phi}\hat{\mathbf{h}}(n-1/n-1) + \mathbf{K}(n)\alpha(n) \tag{16}$$

and

$$\hat{h}(n/n) = \mathbf{g}^T\hat{\mathbf{h}}(n/n) \tag{17}$$

The error covariance matrix is updated as follows:

$$\mathbf{P}(n/n) = \mathbf{P}(n/n-1) - \mathbf{K}(n)\mathbf{s}^T(n)\mathbf{P}(n/n-1) \tag{18}$$

It should be noted that the state vector and the error covariance matrix are initially assigned to zero vector and identity matrix respectively, i.e. $\hat{\mathbf{h}}(0/0) = \mathbf{0}$ and $\mathbf{P}(0/0) = \mathbf{I}_p$.

However, equations (12)-(18) can be carried out providing the AR parameters that are involved in the transition matrix $\boldsymbol{\Phi}$ and the driving process variance σ_u^2 are available. They will be estimated in the next two subsections.

B. Estimation of the AR Parameters

To estimate the AR parameters from the estimated fading process $\hat{h}(n)$, equations (16) and (17) are firstly combined to express the estimated fading process as a function of the AR parameters:

$$\begin{aligned}\hat{h}(n) &= \mathbf{g}^T\boldsymbol{\Phi}\hat{\mathbf{h}}(n-1) + \mathbf{g}^T\mathbf{K}(n)\alpha(n) \\ &= \hat{\mathbf{h}}^T(n-1)\mathbf{a}(n) + \upsilon(n)\end{aligned} \tag{19}$$

where $\hat{\mathbf{h}}(n-1) = [\hat{h}(n-1) \quad \hat{h}(n-2) \quad \cdots \quad \hat{h}(n-p)]^T$ and $\mathbf{a}(n) = [-a_1 \quad -a_2 \quad \cdots \quad -a_p]^T$. In addition, the variance of the process $\upsilon(n) = \mathbf{g}^T\mathbf{K}(n)\alpha(n)$ is given by:

$$\sigma_\upsilon^2(n) = \mathbf{g}^T\,\mathbf{K}(n)C(n)\,\mathbf{K}^H(n)\mathbf{g} \tag{20}$$

When the channel is assumed stationary, the AR parameters are time-invariant and satisfy the following relationship:

$$\mathbf{a}(n) = \mathbf{a}(n-1) \tag{21}$$

As (19) and (21) define a state space representation for the estimation of the AR parameters, a second Kalman filter can be used to recursively estimate $\mathbf{a}(n)$ as follows:

$$\hat{\mathbf{a}}(n) = \hat{\mathbf{a}}(n-1) + \mathbf{K}_\mathbf{a}(n)\left(\hat{h}(n) - \hat{\mathbf{h}}^T(n-1)\hat{\mathbf{a}}(n-1)\right) \tag{22}$$

where the Kalman gain $\mathbf{K}_\mathbf{a}(n)$ and the update of the error covariance matrix $\mathbf{P}_\mathbf{a}(n)$ are respectively given by:

$$\mathbf{K}_\mathbf{a}(n) = \mathbf{P}_\mathbf{a}(n-1)\hat{\mathbf{h}}^*(n-1)\left(\hat{\mathbf{h}}^H(n-1)\mathbf{P}_\mathbf{a}(n-1)\hat{\mathbf{h}}(n-1) + \sigma_\upsilon^2(n)\right)^{-1} \tag{23}$$

$$\mathbf{P}_\mathbf{a}(n) = \mathbf{P}_\mathbf{a}(n-1) - \mathbf{K}_\mathbf{a}(n)\hat{\mathbf{h}}^T(n-1)\mathbf{P}_\mathbf{a}(n-1) \tag{24}$$

with initial conditions $\hat{\mathbf{a}}(0) = \mathbf{0}$ and $\mathbf{P}_\mathbf{a}(0) = \mathbf{I}_p$.

It should be noted that according to (20) and (23), the variance of the innovation process of the first Kalman filter is used to drive the Kalman gain of the second.

C. Estimation of the Driving Process Variance

To estimate the driving process variance σ_u^2, the Riccati equation is first obtained by inserting (14) in (18) as follows:

$$\mathbf{P}(n/n) = \boldsymbol{\Phi}\mathbf{P}(n-1/n-1)\boldsymbol{\Phi}^H + \mathbf{g}\sigma_u^2\mathbf{g}^T - \mathbf{K}(n)\mathbf{s}^T(n)\mathbf{P}(n/n-1) \tag{25}$$

Taking into account that $\mathbf{P}(n/n-1)$ is a symmetric Hermitian matrix, one can rewrite (15) in the following manner:

$$\mathbf{b}^T(n)\mathbf{P}(n/n-1) = C(n)\mathbf{K}^H(n) \tag{26}$$

By combining (25) and (26), σ_u^2 can be expressed as follows:

$$\sigma_u^2 = \mathbf{f}[\mathbf{P}(n/n) - \boldsymbol{\Phi}\mathbf{P}(n-1/n-1)\boldsymbol{\Phi}^H + \mathbf{K}(n)C(n)\mathbf{K}^H(n)]\,\mathbf{f}^T \tag{27}$$

where $\mathbf{f} = [\mathbf{g}^T\mathbf{g}]^{-1}\mathbf{g}^T = \mathbf{g}^T$ is the pseudo-inverse of $\mathbf{g}$.

Thus, we propose to estimate σ_u^2 recursively as follows:

$$\begin{aligned}\hat{\sigma}_u^2(n) = \lambda\hat{\sigma}_u^2(n-1) + (1-\lambda)\mathbf{f}[&\mathbf{P}(n/n) - \boldsymbol{\Phi}\mathbf{P}(n-1/n-1)\boldsymbol{\Phi}^H \\ &+ \mathbf{K}(n)|\alpha(n)|^2\mathbf{K}^H(n)]\,\mathbf{f}^T\end{aligned} \tag{28}$$

where the variance of the innovation process $C(n)$ is replaced by its instantaneous value $|\alpha(n)|^2$ and λ is the forgetting factor. It should be noted that λ can be either constant or time-varying (e.g., $\lambda(n) = (n-1)/n$).

D. Operation of the Channel Estimator

During the so-called training mode, the first Kalman filter in Fig. 1 uses the training sequence $s_m(n)$, the observation $y_m(n)$ and the latest estimated AR parameters $\{\hat{a}_i\}_{i=1,\dots,p}$ to estimate the fading process $h_m(n)$; while the second Kalman filter uses the estimated fading process $\hat{h}_m(n)$ to update the AR parameters. At the end of the training period, the receiver stores the estimated AR parameters and uses them in conjunction with the observation $y_m(n)$ and the decision $\hat{s}_m(n)$ to predict $h_m(n+1)$ in a decision directed manner. It should be noted that a prediction version of the Kalman filtering algorithm (12)-(18) must be used in the decision directed mode.

At the receiver, the received signal is multiplied by the conjugate of the channel estimate to compensate for the phase offset introduced by the fading channel, and the data symbols are recovered by coherent detection.

IV. Simulation Results

A. Simulation Protocols

In this section, we carry out a comparative simulation study on the estimation of OFDM fading channels between the proposed two-cross-coupled Kalman filter based channel estimator and other estimators based on LMS or RLS algorithms [6]. We consider an OFDM system with QPSK modulation, 52 carriers, and a central carrier frequency of 1900 MHz. The transmitted frame size over each carrier is assumed to be 256 symbols. The fading processes $\{h_m(n)\}_{m=1,\ldots,M}$ are generated according to the modified Jakes model [22] with 16 distinct oscillators and Doppler rate $f_d T_s = 0.097$. They are normalized to have a unit variance, i.e. $\sigma_h^2 = 1$. The average Signal-to-Noise Ratio (SNR) per carrier is defined by:

$$\text{SNR} = 10\log_{10}(\sigma_h^2/\sigma_w^2) = 10\log_{10}(1/\sigma_w^2) \quad (29)$$

B. Results and Comments

Fig. 5 illustrates the on-line estimation of the AR(2) parameters, for high Doppler rate scenario of $f_d T_b = 0.097$ and SNR=30 dB. One can notice that the estimated real and imaginary parts of the AR(2) parameters converge to the true values after approximately 100 symbols.

Fig. 6 and 7 show, respectively, the Mean Square Error (MSE) of the estimated fading process and the BER performance of the OFDM system when using the various channel estimators. Thus, exploiting the channel statistics by using AR models in the proposed two-cross-coupled Kalman filter based channel estimator results in significant performance improvement over the LMS and RLS based ones. In addition, increasing the AR model order will improve the performance of the system with the amount of improvement decreases as the model order increases. While the amount of improvement between AR(1) and AR(2) is significant, the amount of improvement beyond AR(5) is not so much. Although high-order AR models (e.g., AR(20)) can provide better modeling approximation than low-order AR models (see Fig. 3 and 4), the amount of performance improvement in that case is small compared with the resulting computational cost $O(p^3)$ of the estimation algorithm. Therefore, to reduce the computational cost, an AR(5) is recommended.

V. Conclusion

This paper investigates the estimation of rapidly time-varying OFDM fading channels. A scheme based on two-cross-coupled Kalman filters is proposed for the joint estimation of the fading process and the corresponding AR parameters over each carrier. The comparative simulation study we have carried out with the conventional LMS and RLS channel estimators shows that the two-cross-coupled Kalman filter based channel estimator can provide significant results over the LMS and RLS ones. In addition, an AR(5) model can provide a trade-off between the accuracy of the model, the computational cost of the estimation algorithm and the system performance.

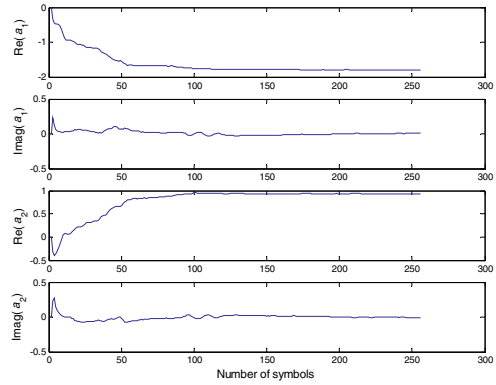


Fig. 5. Real and imaginary parts of the estimated AR(2) parameters of the fading process along the first carrier. True AR(2) parameter values are $a_1 = -1.7627$ and $a_2 = 0.9503$.

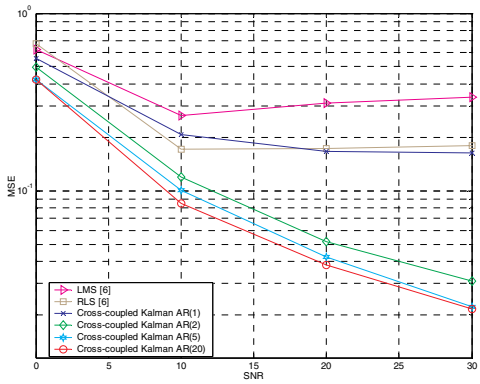


Fig. 6. MSE versus SNR of the OFDM system with the various channel estimators.

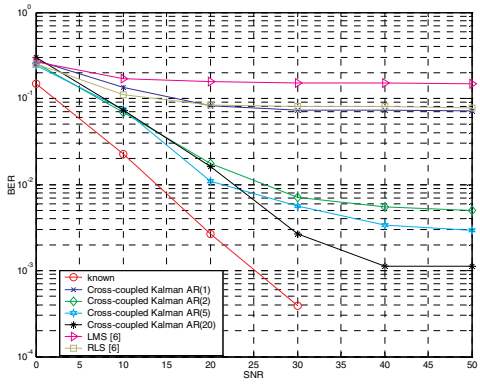


Fig. 7. BER performance versus SNR of the OFDM system with the various channel estimators.

REFERENCES

[1] L. J. Cimini, "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. Commun.*, vol. COM-33, pp. 665-675, July 1985.

[2] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications, where Fourier meets Shannon," *IEEE Signal Processing Magazine*, vol. 17, pp. 29-48, May 2000.

[3] L. Hanzo, M. Munster, B. J. Choi and T. Keller, *OFDM and MC-CDMA for Broadband Multi-user Communications, WLANs and Broadcasting*. Wiley, 2003.

[4] J. K. Tugnait, Lang Tong and Zhi Ding, "Single-user channel estimation and equalization," *IEEE Signal Processing Magazine*, vol. 17, no. 3, pp. 17-28, May 2000.

[5] W. C. Jakes*, Microwave Mobile Communications*. New York: Wiley, 1974.

[6] D. N. Kalofonos, M. Stojanovic and J. G. Proakis, "Performance of adaptive MC-CDMA detectors in rapidly fading Rayleigh channels," *IEEE Trans. On Wireless Commun.*, vol. 2, no. 2, pp. 229-239, March 2003.

[7] C. Komninakis, C. Fragouli, A. H. Sayed and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. On Signal Processing*, vol. 50, no.5, pp. 1065-1076, May 2002.

[8] W. Chen and R. Zhang, **"**Estimation of time and frequency selective channels in OFDM systems: a Kalman filter structure**,"** in *Proc. IEEE GLOBECOM'04*, pp. 800-803, Nov. 2004.

[9] S. B. Bulumulla, S. A. Kassam, and S. S. Venkatesh,"An adaptive diversity receiver for OFDM in fading channels" in *Proc. IEEE ICC'98,* vol. 3, pp. 1325 - 1329, 7-11 June 1998.

[10] L. Lindbom, A. Ahlen, M. Sternad and M. Falkenstrom, "Tracking of time-varying mobile radio channels. II. A case study," *IEEE Trans. On Communications,* vol. 50, no. 1, pp. 156-167, Jan. 2002.

[11] C. Tepedelenlioglu, A. Abdi, G. B. Giannakis and M. Kaveh, "Estimation of Doppler spread and signal strength in mobile communications with applications to handoff and adaptive transmission," *Wireless Commun. Mobile Comput.,* vol. 1, pp. 221-242, June 2001.

[12] M. Tsatsanis, G. B. Giannakis and G. Zhou, "Estimation and equalization of fading channels with random coefficients," *Signal Process.,* vol. 53, pp. 211-229, Sept. 1996.

[13] J. Cai, X. Shen, and J. W. Mark, "Robust channel estimation for OFDM wireless communication systems - an H_∞ approach," *IEEE Trans. On Wireless Commun.*, vol. 3, no. 6, pp. 2060 - 2071, Nov. 2004.

[14] M. Deriche, "AR parameter estimation from noisy data using the EM algorithm," in *Proc. of the IEEE-ICASSP,* Adelaide, April 1994, pp. 69-72.

[15] B. D. O. Anderson and J. B. Moore, *Optimal Filtering.* Englewood Cliffs, NJ: Prentice-Hall, 1979.

[16] S. Gannot and M. Moonen, "On the application of the unscented Kalman filter to speech processing," in *Proc. of IWAENC,* Kyoto, Japan, September 2003.

[17] D. Labarre, E. Grivel, Y. Berthoumieu, E. Todini and M. Najim, "Consistent estimation of autoregressive parameters from noisy observations based on two interacting Kalman filters," *Signal Process.,* vol. 86, pp. 2863-2876, October 2006.

[18] A. Jamoos, D. Labarre, E. Grivel and M. Najim, "Two cross coupled Kalman filters for joint estimation of MC-DS-CDMA fading channels and their corresponding autoregressive parameters," in *Proc. EUSIPCO'05*, Antalya, 4-8 Sept. 2005.

[19] K. E. Baddour and N. C. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Trans. On Wireless Commun.*, vol. 4, no. 4, pp. 1650-1662, July 2005.

[20] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 2002.

[21] D. Schafhuber, G. Matz, and F. Hlawatsch, "Simulation of wideband mobile radio channels using subsampled ARMA models and multistage interpolation," in *Proceedings of IEEE-SSP*, (Singapore), pp. 571–574, Aug. 6–8, 2001.

[22] P. Dent, G. Bottomley and T. Croft, "Jakes fading model revisited," *IEE Electronics Letters*, vol. 29, no. 13, pp.1162-1163, June 1993.

EcoLocate: A Heterogeneous Wireless Network System for Wildlife Tracking

Andrew C. Markham, Andrew J. Wilkinson
Department of Electrical Engineering
University of Cape Town
Cape Town, South Africa
mrkand007@uct.ac.za

Abstract— **All research to date using wireless networks for wildlife tracking has concentrated on monitoring a single species, using large GPS enabled collars. These collars are too heavy to attach on smaller animals. Rather than omit small animals from the tracking spectrum, we show how a fusion of GPS tracking (where applicable) and an improved version of VHF tracking can result in a system which is able to track a wide range of animal species using the same underlying wireless network for information transfer. Tags are equipped with radio transceivers, which are used to both transmit and receive beacons. Received beacons are used to construct radio proximity maps which characterize co-location of various animals at different points in time. Furthermore, as the locations of some nodes are known, coarse estimates of animal locations can be determined, especially around focal points such as waterholes. We present the components of our system and discuss our prototype implementation.**

Keywords-Wildlife tracking; GPS; VHF; Wireless Network; beacon; proximity detection

I. INTRODUCTION

No research to date has considered the incredible diversity in the Animal Kingdom when designing a wireless network for wildlife monitoring and tracking. The preeminent and frequently cited example of wildlife wireless sensor networks is ZebraNet [1], [2]. As the name of the project suggests, zebras were equipped with solar powered GPS tracking collars which transferred information by using bi-directional wireless links to form a network. Information was transferred using a simple flooding protocol such that all collars would share the same data. The motivation for ZebraNet is that by just encountering one or two animals, data from all the other animals can be retrieved, greatly reducing the logistics involved in locating a large number of animals in order to perform a complete download. Whilst their pioneering work provided some valuable lessons, especially regarding the difficulties involved in real world deployment, they only considered one single species of animal. Although there is nothing preventing the same ZebraNet collars from being placed on similar sized or larger animals (such as buffalo or elephant), smaller mammals cannot be equipped with such heavy and large collars. This is due to guidelines from the wildlife tracking community, who recommend that tag weight should not exceed 3-5% of animal bodyweight [3]. Thus, a valuable cross-section of the Animal Kingdom is ignored or omitted, whereas zoological researchers need to be able to monitor as wide a range as possible. From multi-species data, they can infer both inter- and intra- specific behaviour, and propose guidelines for wildlife management. However, the power consumption of current GPS receivers precludes their use on smaller animals. Although the receiver itself is light (in the order of ten grams) the battery capacity required to keep the unit operating for a significant monitoring duration (a year or more) is large (a few hundred grams).

However, this does not mean that smaller wildlife should be omitted entirely from the spectrum of wildlife tracking. Merely because GPS technology cannot be used, this does not mean that no technology at all should be used. Rather, we turn to the original technology that was placed on wildlife before the advent of satellite location, namely VHF tracking and triangulation.

This paper first reviews conventional VHF tracking illustrating some of the difficulties of this technique. Next, we discuss how the concepts behind VHF tracking can be improved by equipping tags with bi-directional radio-transceivers. The components of our system are then introduced and power consumption and lifetime are discussed. Finally, we compare similar work, pose future directions and draw conclusions.

II. CONVENTIONAL VHF TRACKING

VHF tracking involves placing small RF transmitters on animals [4]. These emit a signal every few seconds on a dedicated narrow frequency band. Researchers then estimate the location of an animal by obtaining bearings to the animal from two or more known locations using directional Yagi antenna and a handheld receiver. The strength of the signal is typically gauged manually using headphones, and the bearing is obtained by rotating the antenna in the direction of maximal loudness. Obtaining location estimates is a laborious and time consuming process, especially if large numbers of animals are required to be located [3], [4]. Furthermore, location estimates have accuracies in the range of tens to hundreds of metres as opposed to the impressive (< 10 [m] CEP [5]) accuracy of GPS receivers. Even though the estimates are much less accurate, the simplicity and small size (a few grams) of VHF beacons

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 293–298.

make them an attractive (and in most cases, the only) option for smaller animals.

The goal of our research is to provide a framework in which VHF tracking and GPS receivers are combined with the increased functionality provided by the wireless network. Large animals are equipped with full function GPS tracking collars with radio transceivers for two way communication, whilst the smaller animals are just equipped with the lightweight radio transceivers. Operation is controlled by a microcontroller and collars are powered by a battery (augmented in some cases with a solar panel). In the contemporary VHF scenario, the collars have a transmit-only radio unit, but we show how beacons can be used to obtain coarse positional or contextual information through receiving other beacons within radio range.

III. Beacons for Proximity Detection

Each tag[1] in the network periodically emits a beacon signal. Where this differs from conventional VHF tracking is that the beacon signal is a digital packet as opposed to a single pulse. Amongst other data, the transmitted packet of data contains the unique ID (identification) code of the tag. Another difference with existing VHF technology is that all the tags transmit on the same channel. Nodes access the channel randomly, which obviates difficulties inherent in time synchronizing a sparse, mobile network, but also results in periodic packet loss through collision, when two or more tags attempt to access the medium simultaneously. However, the average time interval between beacon transmissions is long (typically 3 s) compared to the time required to transmit a packet (6 ms). Furthermore, as the network is extremely sparse, the number of nodes in close radio proximity is typically very small resulting in a low probability of collision.

In standard VHF tracking, there are many transmitting beacons and a single receiver. However, as tags in our system are equipped with a bi-directional radio transceiver, tags can both transmit and receive beacons. Thus, a tag periodically listens to the beacon channel for a length of time and captures the ID numbers of the nodes within radio proximity. As all the beacons transmit on the same channel, a tag does not have to scan through multiple frequencies, reducing the overall scan time. Nodes store the identities of the beacons that were overheard within the listening window, along with a time-stamp indicating time of reception. Typical activity on the beacon channel is illustrated in Fig. 1. This data is then transferred through the multi-hop wireless network to ultimately reach a base-station, where a logical connectivity map can be constructed. The connectivity map is a time-varying representation of the radio proximity of various tags. Based on this data, inferences can be made about the habits and behaviours of animals.

This approach dramatically increases the usefulness of the tracking system, as the map essentially characterizes the co-location of tags, albeit coarsely. This information can be used by researchers to determine if a certain animal is frequently in contact with other particular animals and also captures the time

[1] The terms tag, collar and node are used interchangeably in this paper

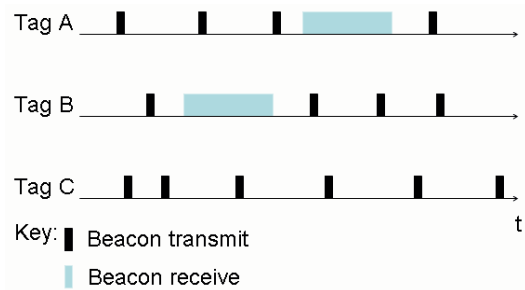


Figure 1. Timeline showing tag activity in the beacon channel. When tags are not transmitting nor receiving, they enter a low power sleep mode to conserve power.

dependency of this relationship, which can provide valuable insight into animal behaviour. For example, some animals, such as leopards, are solitary except during the mating season. By tagging these creatures, and monitoring their proximity, the genetic lineage can be discovered which provides invaluable information to the field of wildlife research [6]. Furthermore, the proximity of animals to human settlements or park boundaries can be monitored which can provide an early warning about rogue or problem animals.

However, this is not all that the time varying connectivity map can tell. If the locations of some nodes are known (either in fixed locations or mobile and equipped with GPS receivers), the approximate location of an animal not carrying a GPS receiver can be determined. With an increasing number of simultaneous location estimates, the accuracy of the predicted position can be refined. Such location estimates, whilst very coarse in relation to GPS fixes, can provide important information about the times at which animals visit certain locations such as waterholes and how long they remain in that area. Thus, this moves from simple (and laborious) VHF tracking to a system which can generate a wealth of interesting and useful data. The ramifications of such an approach are not just directed towards the field of wildlife research, but also to more general aspects of wildlife management. Thus decisions about the number and placement of waterholes, size and range of game reserves, and allocation of resources can be performed in a scientific and informed manner.

The system components which are used to acquire and transfer this data are now introduced.

IV. System Components

As mentioned previously, the highly diverse Animal Kingdom demands a widely heterogeneous solution. To this end, the wireless network is comprised of four different classes of tags, with increasing levels of functionality. In addition to increased features, the higher level nodes have the same functionality of all the lower level nodes. This leads to a unified design approach, as they all run the same basic firmware except certain functions are disabled based on their class. The development cycle is consequently simplified as only one source code has to be maintained – the class of the tag is chosen either at design time (through conditional compiler switches) or dynamically at run time in the field (through

assessment of their relative rank in the system [7]). The tags in increasing order of functionality (and correspondingly increasing order of power consumption and weight) have been termed Marker nodes, Spotter nodes, Pack nodes and Base-station nodes.

In our prototype implementation we have used the Microchip PIC18LF series of microcontrollers [8] and the Nordic NRF905 UHF radio transceiver [9] operating in the 868 MHz ISM license free (European and African) band. Tag operation is controlled using a finite state machine based operating system, where the choice of available states is controlled by the class of the tag. The radio transceiver achieves a quoted transmission range of 300 m at +10 dBm output power [9], but tests in the field show that the range is more typically 150 to 200 m. Depending on local topographic features, the radio transmission range can drop to as low as 30 m. For GPS enabled nodes, the u-blox LEA-4P receiver is used to determine location [10]. Information is stored either in the FLASH memory of the microcontroller itself, or using an SD card which provides an inexpensive method of storing large (512 Mb or more) volumes of data. Some nodes have also been equipped with solar panels to provide a renewable source of energy. Figures regarding power consumption are thus based on these set of components, assuming a nominal 3V power source. Typical values for current consumption, as well as predicted lifetime based on battery with 1 Ah of capacity are shown in Table 1.

TABLE I. POWER CONSUMPTION OF THE VARIOUS CLASSES OF TAGS WITHIN THE ECOLOCATE SYSTEM

Tag Type	Operation	Duty cycle	Current (mA)	Mean Current (mA)	Lifetime per Ah (days)
Marker	Sleep	99.8%	0.04	0.11	386
	Beacon Tx	0.2%	34		
Spotter	Sleep	99.2%	0.04	0.21	200
	Beacon Tx	0.2%	34		
	Beacon Rx	0.5%	15		
	Upload	0.1%	26		
Pack	Sleep	98.3%	0.04	0.44	94
	Beacon Tx	0.2%	34		
	Beacon Rx	0.5%	15		
	Route	1.0%	26		
GPS Enabled Pack	Sleep	95.0%	0.04	1.71	24
	Beacon Tx	0.2%	34		
	Beacon Rx	0.5%	15		
	Route	1.0%	26		
	GPS fix	3.3%	38		

A. *Marker Tags*

Marker tags are the simplest type of tag in the network and have extremely low power consumption. Thus, this means that they can be fitted with a lightweight battery, making them suitable for tagging small animals. These tags just act as beacons, periodically emitting ('marking') their ID and other salient information, such as temperature and movement parameters in a packet. Marker tags never listen for other node's IDs and consequently consume a miniscule amount of power as they spend the majority of their time in low power sleep mode. Tags randomly wake up to transmit a packet before returning to sleep mode. Using the prototype node, if a node transmits its ID on average every three seconds, it will last for approximately 380 days per Ah of battery capacity. With two small coin cell batteries (such as CR-2032 [11]), a year of life can be obtained from a tag with total weight of 15 g.

Two unpackaged Marker tags are shown in Fig. 2, demonstrating their small size. In this picture, the antenna has not yet been attached. The 28 pin 18LF2620 microcontroller is used for the Marker tag and is wired up to the transceiver board. Although labour intensive, this results in a lower overall tag weight.

B. *Spotter Tags*

These tags provide all the functionality of the Marker nodes but periodically 'spot' or listen to other nodes within their radio range. They store the overheard node IDs and any other transmitted data in memory, along with a timestamp. When in range of a pack or base-station tag they transfer the stored information through wireless network. However, they only act as leaf or end nodes in the network and do not route other nodes' packets. Their power consumption is higher than the simpler Marker type tags, as they have to remain awake in active receive mode for twice the average inter-beacon time due to the unsynchronized nature of the channel access. Thus the receive window is six seconds long if the average inter-beacon transmission time is three seconds. For a Spotter tag which transmits a beacon every three seconds and listens for beacons every ten minutes, it will survive for approximately 280 days per Ah of battery capacity. Thus, a tag powered by two 'AA' sized batteries would last for over a year and weigh approximately 40 g [11].

Three Spotter tags are shown in Fig. 3. These are powered by a single lithium CR123A [11] battery which provides over six months of operation. They have been encapsulated in quickset epoxy to make them robust and waterproof. The tags are fitted with a quarter wave antenna fabricated from a length of 0.5 mm steel wire. Proximity logs are stored on-board in the 64 kbyte FLASH memory of the microcontroller, ready for upload to Pack tags when within range.

C. *Pack tags*

Pack tags form the multi-hop network. They perform all the tasks of Spotter nodes and also route information through the wireless network in a store-and-forward fashion. As the radio radius is small compared to the total area the network is very sparse, thus information is transferred opportunistically upon contact with another Pack tag or base-station.

To decide how to route packets, nodes assess their ranking (the ranking is similar to reputation schemes in wireless security) in terms of the global distribution of resources (such as remaining battery energy or connectivity) using local information. This is dynamically assessed, and thus if new nodes (with large amounts of battery energy) are inserted into the network, these new nodes will assume a higher rank and participate more fully in the processes of routing, removing the load of routing from nearly exhausted nodes. Furthermore, as a node nears the end of its lifetime it can remove certain functionalities from the finite state machine, becoming a spotter node and ultimately a marker node. In this way, data can still be obtained from a tagged animal for much longer than if its tag was always providing a large amount of functionality.

Figure 2. Two unpackaged marker tags. The microcontroller (18LF2620) can be seen on the left tag and the radio transceiver on the right tag. An 'AA' battery is shown for scale purposes.

The details of determining a node's rank within the network are beyond the scope of this paper and can be found in [7] which discusses a single copy routing strategy and [12] which extends the framework to a multi-copy routing strategy with replication and redundancy.

Pack tags transmit a special beacon packet which contains information which other tags use to update their ranking within the network as well as their ID. Thus, the beacon packet acts both as a proximity locator for the purpose of animal tracking and also as a network control or discovery packet. As the NRF905 is frequency agile, a pack tag advertises a random channel to use for network data transfer when it sends the beacon packet. This channel is separate from the beacon channel, so the act of transferring data does not affect beacon transmissions at all – in fact, the networking and the beaconing are completely distinct from one another.

Quantifying energy usage of pack tags is difficult, because it depends on a variety of factors such as the network density, volume of traffic and network composition. However, typical lifetimes are in the range of 90 days per Ah of battery capacity. Nodes thus need two alkaline 'C' cells to survive for a year or more, resulting in a tag weight of approximately 150 g [11]. If the tag is also equipped with a GPS receiver (which takes fixes every 15 minutes with an average fix time of 30 seconds) it requires six 'C' cells, resulting in a tag weight of approximately 400 g [11]. This example also demonstrates the high power consumption of the GPS receiver and why it is unsuitable for use on smaller animals. This weight can be reduced significantly if the tag is equipped with a solar panel.

Fig. 4 shows a Pack tag. This is an early prototype version. It is equipped with a PIC18LF4620 microcontroller, a GPS receiver and a 512 Mb SD card for data storage. In addition to the standard Pack tag features, it also has a tri-axial accelerometer used for measuring animal motion and scheduling GPS fixes. A temperature sensor is used to monitor the ambient temperature, and a photo-sensor measures the incident light level. Our new version of this tag is greatly reduced in size and has the ability to recharge the battery using a flexible solar cell.

Figure 3. Three spotter class tags. The units have been packaged in epoxy resin for strength and waterproofing. The tags are equipped with a rigid quarter wave whip aerial. An 'AA' battery is shown for scale purposes.

D. Base-station Tags

These nodes act as sink nodes in the network and provide an interface between the wireless data gathering network and the end users. They are essentially pack tags except that all received data is forwarded out of the network via some communications interface. This interface can either be wired or wireless, depending on the application. Base-stations can be placed at convenient sites (such as at the top of a hill) or at points of attraction for animals (such as waterholes or salt licks). Base-stations can be mobile, and these can be carried by people (game-rangers or tourists), attached to vehicles or affixed to large animals, such as elephants. The choice of the communications interface depends on the local infrastructure, but can be cellular GSM modem, UHF modem or even satellite upload. As the communications interface is long range, its power consumption is typically large. Thus, base-station tags should be equipped with a solar panel or attached to a source of mains power if possible.

Base-station nodes can also be equipped with directional antennas. This will increase the accuracy of location estimates, as the bearing to a beaconing node can be determined. In addition, fixed base-stations can be equipped with more sensitive radio receivers and larger antenna, which will increase the range at which beacons can be detected. Thus, many of the well established techniques of conventional VHF tracking can be carried over to our system.

A base-station node is shown in Fig. 5. This node provides an interface between the wireless animal tracking network and a PC using a serial link. The base-station can also act in passive listening mode, where it continually listens to beacons and reports overheard beacons to the host program.

V. Related Work

ZebraNet provided a comprehensive test on the use of wireless sensor networks for animal tracking [1], [2]. GPS equipped collars were fitted on zebras and exchange information in an epidemic fashion. Their routing algorithm is

Figure 5. An early prototype pack tag. This tag is equipped with a GPS unit, tri-axial accelerometer, light and temperature sensors.

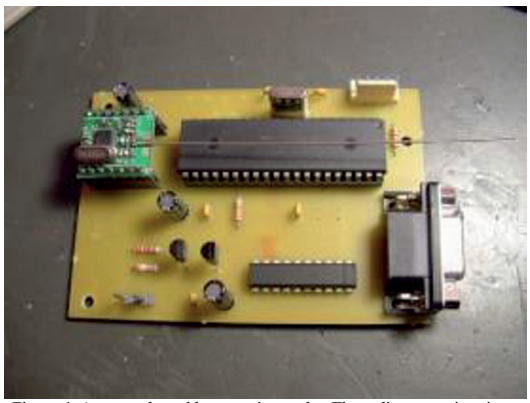


Figure 4. An unpackaged base-station node. The radio transceiver is on the left, a PIC18LF4620 microcontroller is to the right. The unit is connected to a PC via an RS-232 link.

very simple and leads to buffer overflow as every node in the network stores information from every other node, thus limiting scalability. The authors only considered fitting the collars on a single type of animal. The Shared Wireless Infostation Model (SWIM) is a routing protocol that addresses some of the issues faced by the Epidemic routing protocol [13]. Their main contribution was in the form of 'anti-packets' – messages that prevent nodes from buffering data that has already been delivered to the base-station. However, like ZebraNet, they concern themselves with instrumenting a single species – whales.

Sikka et al. present a wireless sensor network designed to monitor a typical farm environment, in particular cattle monitoring [14], [15]. Using the capabilities of different animals to lead to a better performing network is also not considered in their work. In addition, an agricultural deployment is much simpler as collars can be easily replaced when they fail. A similar agricultural monitoring system is discussed in [16]. Likewise, they only concern themselves with instrumenting one type of animal.

A similar scheme to ZebraNet is CenWits, which is a wireless network based Search-and-Rescue system. Nodes are equipped with GPS receivers and exchange databases of recent locations of other nodes when they meet. Based on this data, a rescue team can determine the location where a missing person was last in contact with another node [17].

Proximity detection between groups of animals has recently been undertaken using commercially available collars [18]. Like other research this only concentrated on a single animal species, namely raccoons. These collars do not form a wireless network and have to be retrieved in order to download data from them using a cable. This severely limits their usefulness for wide ranging species, and places additional stress on animals as they have to be recaptured.

To the best of our knowledge, our system is the first work which has combined the simplicity of VHF tracking and the accuracy of GPS, along with the flexibility of a wireless network communications overlay in order to track multiple species using the same basic technology.

VI. Future Work

This work is at an advanced stage of development, with working prototype tags having been constructed. Three Spotter class tags and a base-station tag have been deployed at Ongava Research Centre in Namibia in order to test their operation in the field. The Spotter tags have been attached to vehicles, and have been operating correctly for the past three months. The next stage of the project is to deploy a small test system on wild animals. Because of the high cost of deployment (helicopter, veterinarians and tranquilizer drugs), our tags will be piggybacked onto existing commercial collars for preliminary trials. An advantage of this approach is that our retrieved data can be compared with that of the commercial collar.

The system is not only suitable for tracking animals, but can also be used to monitor a large variety of factors which influence animal behaviour. For example, water availability plays a large role in animal location. Waterholes and rivers can be equipped with water-level sensors which can relay information about the amount of water in an area through the wireless network to a researcher. Based on park-wide information on water distribution, informed decisions can be made about where to provide more sources of water. Other sensors can also be installed, such as fire detectors, soil moisture and temperature sensors amongst others. We have developed a water level sensor but have yet to test it in the field.

An important issue is field reprogramming. In order to provide firmware updates and fix errors without having to retrieve collars (which in most cases is difficult or impossible), it is desirable to be able to program the collars ‘over-the-air’. We are planning to implement this feature in our next version.

VII. Conclusions

We have presented a new system for monitoring multiple species using both GPS tracking where possible and techniques from conventional VHF tracking. Simple tags carried by small

animals use the increased capabilities of the large tags carried on larger animals to carry data in an energy efficient manner to the end user. Information is transferred using a delay tolerant network where contacts are formed opportunistically when tags are within range of each other. By using a mixture of technologies, a wide range of different animal species can be simultaneously tracked, using the same basic hardware. This work has the potential to revolutionize the field of wildlife tracking, by unifying two currently disparate tracking technologies and meshing them with the power of a mobile wireless network.

ACKNOWLEDGEMENTS

We thank Dr Ken Stratford for his helpful comments and for deploying the tag devices in the field. Andrew Markham is supported by a UCT International Students Scholarship and thanks the Postgraduate Funding Office for their assistance.

REFERENCES

[1] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. Peh and D. Rubenstein, "Energy-Efficient Computing for Wildlife Tracking: Design Tradeoffs and Early Experiences with ZebraNet", *ASPLOS X*, San Jose, CA, USA, 2002

[2] P. Zhang, C. M. Sadler, S. A. Lyon and M. Martonosi, "Hardware Design Experiences in ZebraNet", *SenSys*, Baltimore, Maryland, USA, 2004

[3] W. Cochran, "Wildlife Telemetry", In S. D. Schemnitz (Ed.), *Wildlife management techniques manual*, 1980

[4] R. E. Kenward, *A Manual for Wildlife Radio Tagging*, Academic Press, 2001

[5] Kaplan (Ed.), *Understanding GPS: Principles and Applications*, Artech House, 1996

[6] G. Spong, M. Johansson and M. Björklund, "High genetic variation in leopards indicates large and long-term stable effective population size", *Molecular Ecology*, vol. 9, no. 11, 2000 pp. 1773

[7] A. C. Markham and A. J. Wilkinson, "The Adaptive Social Hierarchy - A Self Organizing Network Based on Naturally Occurring Structures", In *Proceedings of the First International Conference on Biologically Inspired Networks and Computation (BIONETICS)*, Cavelese, Italy, 2006

[8] Microchip Technology, *http://www.microchip.com*

[9] Nordic Semiconductor, *http://www.nordicsemi.com*

[10] u-blox, *http://www.u-blox.com*

[11] Sanyo battery specifications, *http://us.sanyo.com/batteries*

[12] A. C. Markham and A. J. Wilkinson, "A biomimetic ranking system for mobile energy constrained wireless networks", In *Proceedings of the Tenth Southern African Telecommunications, Networks and Applications Conference (SATNAC)*, Sugar Beach Resort, Mauritius, 2007

[13] T. Small, Z. J. Haas, A. Purgue and K. Fristrup "The Shared Wireless Infostation Model: A New Ad Hoc Networking Paradigm (or Where There is a Whale, There is a Way)", *Fourth ACM International Symposium on Mobile AdHoc Networking & Computing*, 2003

[14] P. Sikka, P. Corke, P. Valencia, C. Crossman, D. Swain and G. Bishop-Hurley, "Wireless adhoc sensor and actuator networks on the farm", In *Proceedings of the Fifth International Conference on Information Processing in Sensor Networks*, Nashville, Tennessee, USA, April 2006

[15] T. Wark, P. Corke, P. Sikka, L. Klingbeil, Y. Guo, C. Crossman, P. Valencia, D. Swain and G. Bishop-Hurley, "Transforming Agriculture through Pervasive Wireless Sensor Networks", *IEEE Pervasive Computing*, vol. 6, no. 2, pp. 50-57, Apr-Jun, 2007

[16] M. Radenkovic and B. Wietrzyk, "Wireless Mobile Ad-Hoc Sensor Networks for Very Large Scale Cattle Monitoring", In *Proceedings of the 6th International Workshop Applications and Services in Wireless Networks* (ASWN 06), 2006, pp. 47–58

[17] J.H. Huang, S. Amjad and S. Mishra, "CenWits: a sensor-based loosely coupled search and rescue system using witnesses", *In Proceedings of the 3rd international conference on Embedded networked sensor systems (SenSys '05)*, San Diego, California, USA, 2005

[18] S. Prange, T. Jordan, C. Hunter and S D Gehrt, "New Radiocollars for the Detection of Proximity among Individuals", *Wildlife Society Bulletin*, vol. 34, no. 5, Dec 2006, pp. 1333

Enhancement of Throughput in 802.15.4 MAC Layer Using the Principle of Circularity

R Bhakthavathsalam
SERC, Indian Institute of Science; Currently on sabbatical at *Tata Elxsi,* Bangalre 560048, INDIA
Tel: +91 80 22979872 Fax: +91 80 22979441 Email: bhaktha@tataelxsi.co.in

Abstract - **IEEE 802.15.4 is a proposed standard addressing the needs of low-rate wireless personal area networks LR-WPAN with a focus on enabling the wireless sensor networks. This is an emerging application of wireless sensor networking with particular performance constraints, including power consumption, physical size, robustness and security. The two factors that are of utmost importance in the development of such networks are energy efficiency and bandwidth utilization. Greater emphasis is hence required for the design of newer MAC protocols to maximize net throughput for the available bandwidth. In this paper, we introduce a new paradigm of 'circularity' by selectively sensing the medium as busy for appropriate data packets in order to obviate the costly data collisions so as to achieve maximum net throughput. Results are substantiated through NS-2 simulations, dealing with IEEE 802.15.4 beacon enabled mode for star topology network.**

Index Terms — **MAC layer, Stack aware, Slotted CSMA, Circularity, Throughput.**

I. INTRODUCTION

In the last few years there has been a significant increase in the research studies directed towards ad-hoc and sensor wireless networks. Terms such as pervasive computing and smart spaces are being used to describe the future of computing and communication at large. These concepts allude to our personal and business domains being densely populated with miniature sensors, which are constantly monitoring the environment and reporting the data to each other or to some central base station. Likewise, IEEE 802.15.4-"Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (LR-WPANs)" represent a milestone in wireless personal area networks and wireless sensor networks **[1]**. IEEE 802.15.4 is a new standard designed for low rate wireless personal area networks. It targets low data rate, low power consumption and low cost wireless networking and offers device level wireless connectivity **[2]**.

A host of new applications can benefit from the new standard, such as those using sensors for security and health monitoring, wall switches that can be moved at will, wireless computer peripherals, controllers for interactive toys, smart tags and badges, tire pressure monitors in cars, and inventory tracking devices. The main goal of 802.15.4 is to provide a physical and MAC layer standard with ultra low complexity, cost, and power for low data rate wireless connectivity among cheap fixed devices. The IEEE 802.15.4 standard supports multiple network topologies, including both star and peer-to-peer networks depending on the application requirements. In the star topology network, all communications must go through the PAN coordinator. In the peer-to-peer topology, the devices can communicate with one another directly, but still the PAN coordinator must be present. The standard also defines possible channel access mechanisms, depending on whether beacon frames (sent periodically by the PAN coordinator to synchronize communications) are used or not.

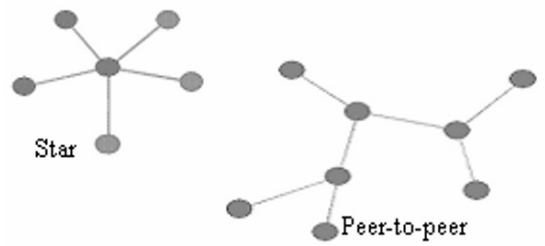


Figure 1: Star and peer-to-peer topology (clustered network)

The challenges in developing such complex network and evaluating its performance metrics are diverse. Only the stack aware features of MAC sublayer for such a wireless network with effective utilization of the available bandwidth are given major attention in this paper. And also we consider beacon-enabled networks, which use slotted carrier sense multiple access mechanism with collision avoidance (CSMA-CA). The non-beacon enabled networks that use simpler unslotted CSMA-CA, similar to the one used in IEEE 802.11 standard will not be considered here **[3]**.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 299–304.

Owing to the scarcity of the available bandwidth, it has become imperative to design a new paradigm such that it uses the bandwidth efficiently. To assess the efficient usage of bandwidth of IEEE 802.15.4 MAC protocol, we introduce a new paradigm of circularity to obviate the costly data collisions in order to maximize net throughput. Section II of this paper presents Stack aware MAC layer. A technical overview of the IEEE 802.15.4 standard is presented in section III. In Section IV a new paradigm is proposed, while the simulation results and concluding remarks are presented in sections V and VI respectively.

II. STACK AWARE *MAC* LAYER

The present OSI reference model has been designed primarily for fixed networks largely comprising of the Internet, which in turn uses the four layer TCP/IP model. Emerging LR-WPAN with a focus on enabling wireless sensor networks have also been modeled based on this layered concept, which however has thrown up some basic challenges. Dynamically changing network topologies and the absence of a sustained source of power in LR-WPAN with increased routing framework requirements have resulted in the need of developing a "Stack Aware" architecture especially for these types of networks. In such architecture, each layer should be aware of what the other layers do and the functionalities in a particular layer are highly optimized for usage with other layers. Physical layer aspects are however not dealt with too deeply. The network and link layers are the two primary layers where effective enhancements need to be made to fully implement the above-mentioned Stack-Awareness for LR-WPAN. Presently, here we concentrate on the Medium Access sublayer of the Data Link layer. It has been seen that the practical capacity of this networks are quite low, largely due to interference and the behavior of the IEEE 802.15.4 Mac Layer protocol **[4]**. As the density of the nodes increases, so does the interference. The contention access period (CAP) and the guaranteed timeslot (GTS) are put in operation here to enable two nodes communicating with each other. However, due to interference and other factors, the data packets may be lost or garbled hence delaying communication between the nodes.

Our goal here is hence to take note of the fact that an increase in the density of the nodes leads to degradation in throughput. An implicit goal in of any MAC protocol is to minimize collisions, raise throughput and prove stable. The two factors, which are of utmost importance in the development of newer MAC protocols, are energy efficiency and bandwidth utilization. Emerging solutions to the Energy problem have been proposed which are independent of the underlying routing protocol and hence they fail to optimize performance for any particular routing framework. Another major factor is the optimization of bandwidth utilization. Bandwidth is a scarce resource and needs to be properly managed if we are to maximize the net throughput. Common solutions include splitting the available bandwidth into the time slots for data packets transmission. The ratio of this division has to be optimal to ensure increased throughput even though this scheme is largely dependent on the channel contention mechanism in use by the protocol. We thus realize that an effective MAC protocol goes a long way in ensuring better network performance, which acts as our primary objective to improve upon the current IEEE 802.15.4 protocol in terms of throughput and thereby achieve effective bandwidth utilization.

III. IEEE 802.15.4 OVERVIEW

A. Physical Layer

There are two physical layer variants in 802.15.4: 868/915MHz and 2.4 GHz **[5]**. The standard defines the 868 MHz band as a single channel with a data rate of 20 kbps, the 915 MHz band as a single 40 kbps channel, while the 2.4 GHz band is divided into 16 channels each with a data rate of 250 kbps. For convenience, this work considers only 2.4 GHz.

B. Medium Access Control Layer

Two topologies are supported by 802.15.4: star and peer-to-peer. Only star networks are considered in this work. The main advantage of using a star network for LR-WPAN applications is that an external coordinator can be used with access to rechargeable power supply. With the star topology there are two communication methods: beacon mode and non-beacon mode. In beacon mode, communication is controlled by the network coordinator, which transmits regular beacons for device synchronization and network association control **[6]**. The network coordinator defines the start and end of a superframe by transmitting a periodic beacon. The length of the beacon period and hence the duty cycle of the system can be defined by the user between certain limits as specified in the standard. The advantage of this mode is that the coordinator can communicate at will with the nodes. The disadvantage is that the nodes must wake up to receive the beacon. In non-beacon mode a network node can send data to the coordinator at will using CSMA/CA if required. However, to receive data from the coordinator, the node must power up and poll the coordinator. To achieve the required node lifetime the polling frequency must be pre-determined by power reserves and expected data quantity. The advantage of non-beacon mode is that the node's receiver does not have to regularly power up to receive the beacon **[7]**.

The disadvantage is that the coordinator cannot communicate at will with the node but must wait to be invited by the node to communicate. In beacon mode, the superframe may consist of both an active and inactive period (Figure 2). The active portion of the superframe, which contains 16 equally spaced slots, is composed of three parts: a beacon, a contention access period (CAP), and a contention free period (CFP). The beacon is transmitted without the use of CSMA at the start of slot 0 and the CAP commences immediately after

the beacon. The coordinator only interacts with nodes during the active period and may sleep during the inactive period. There is a guaranteed timeslot (GTS) option in 802.15.4 to allow lower latency operation. There are a maximum of 7 of the 16 available timeslots that can be allocated to nodes, singly or combined. When a node is allocated a timeslot, it may only transmit data during that timeslot. GTS nodes must listen to the beacon to synchronize prior to communication within its allocated timeslot(s). The relative size of each of the active and inactive periods is determined by the values of the *macBeaconOrder* (*BO*) and the *macSuperframeOrder* (*SO*) and the overall superframe length (or *Beacon Interval*, *BI*) and active superframe duration (*ASD*) are calculated as follows:

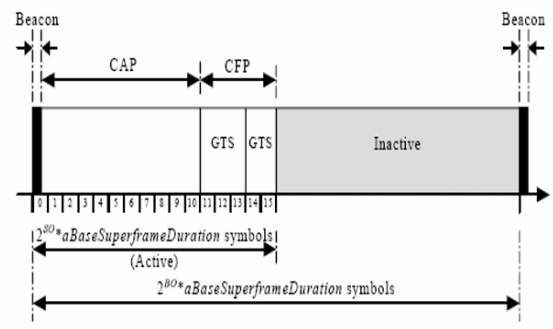


Figure 2: Superframe structure

$$BI = (aBaseSlotDuration \times aNumSuperframeSlots \times 2^{BO})\ symbols$$

$$where \quad 0 \le BO \le 14,\ aBaseSlotDuration = 60\ symbols$$

$$and\ \ aNumSuperframeSlots\ (slots\ in\ a\ superframe) = 16$$

$$ASD = (aBaseSlotDuration \times aNumSuperframeSlots \times 2^{SO})\ symbols$$

$$where \quad 0 \le SO \le BO \le 14$$

The beacons are used to synchronize the attached devices, to identify the PAN, and to describe the structure of superframe. In the slotted CSMA-CA channel access mechanism, the backoff slot boundaries of every device in the PAN are aligned with the superframe slot boundaries of the PAN coordinator. Each time a device wishes to transmit data frames during the CAP, it must locate the boundary of the next slot period. Each device in the network has three variables: NB, CW and BE. NB is the number of times the CSMA-CA algorithm was required to delay while attempting the current transmission **[8]**. It is initialized to 0 before every new transmission. CW is the contention window length, which defines the number of slot periods that need to be clear of activity before the transmission can start. It is initialized to 2 before each transmission attempt and reset to 2 each time the channel is assessed to be busy. BE is the backoff exponent, which is related to how many slot periods a device must wait before attempting to assess the channel. Although the receiver of the device is enabled during the channel assessment portion of this algorithm, the device must discard any frames received during this time. The slotted CSMA-CA mechanism works as follows: NB, CW and BE are initialized and the boundary of the next slot period is located (step1). The MAC layer delays for a random number of complete slot periods in the range 0 to 2^{BE}-1 (step 2) and then requests PHY to perform a CCA (clear channel assessment) (step 3). The MAC sublayer then proceeds provided that the remaining CSMA-CA algorithm steps—frame transmission and any acknowledgement—can be completed before the end of the CAP. If the MAC sublayer cannot proceed, it must wait until the start of the CAP in the next superframe and then repeat the evaluation. If the channel is assessed to be busy (step 4), the MAC sublayer increments both NB and BE by one, ensuring that BE is not more than *aMaxBE*, and CW is reset to 2. If the value of NB is less than or equal to *macMaxCSMABackoffs*, the CSMA-CA must return to step 2, else the CSMA-CA must terminate with a Channel-Access-Failure status. If the channel is assessed to be idle (step 5), the MAC sublayer must ensure that the contention window is expired before starting transmission **[9]**. For this, the MAC sublayer first decrements CW by one. If CW is not equal to 0, it must go to step 3, else start transmission on the boundary of the next slot period.

IV. NEW PARADIGM OF CIRCULARITY

In IEEE 802.15.4, MAC does not suggest how efficiently allocate the ratio of CAP and GTS in the active portion of the superframe to achieve less collision of DATA packets while using the contention access period. Therefore, an efficient design of IEEE 802.15.4 MAC is required for better MAC performance in order to achieve effective utilization of available Bandwidth **[10]**. We propose a new principle called 'Circularity' for LR-WPAN to enhance the MAC performance **[11].**

Our main design objective is to provide less collision, and thereby maximize the net throughput. Circularity concentrates on the modification of DATA transmission from the each node with respect to the time so as to minimize DATA collisions and hence improving the throughput. Here we would like to define circularity as "The total DATA transmissions for each node are divided into specific groups. Each group contains equal number of transmissions. The number of DATA transmissions in each of the groups is termed as circularity. Then each node is assigned a specific circularity value for their DATA packets". Mathematically, a Data packet is circularity-satisfied if:

$$N \text{ MODULO } K = 0$$

where N is the current count of the number of Data packets generated and 'k' is the circularity value for the particular node **[12].**

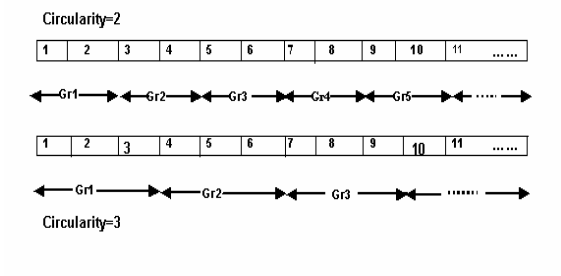


Figure 3: Paradigm of circularity

Now consider the Circularity 2, when the channel is treated as busy even though it is idle from the corresponding node for its second DATA packet transmission, and then select another random number with respect to backoff window. The numbers 1 to 11 and so on refer to the DATA packet number transmitted by that particular SS. The above discussed circularity concept is used to treat the channel as busy selectively for particular DATA packets, irrespective of the consequences. There arise a question that how this sensing mechanism for particular DATA packets irrespective of the network conditions result in improving the MAC performance of the IEEE 802.15.4. Careful analysis and observation proved that the resulting consequences prove to be beneficiary. By treating the channel as busy for that particular DATA packet, we are refraining from initiating a data transfer, which may later prove to be costly if a DATA collision occurs due to contention. When a DATA packet collision occurs, the whole contention resolution phase is repeated. This leads to increase in access delay, which has the impact on the net throughput and necessity of CAP. Treating channel as busy avoids such situations and the concept of circularity is implemented for the above-mentioned reasons. With optimum value of circularity for low and high density of nodes in a network scenario lesser number of DATA packets are refraining from transmission thereby making the system coverage a steady state. Treating channel as busy with irrespective of the consequences for particular DATA packets can be explained through the flow chart as shown.

An integer value is given to circularity and the value of counter is made to vary from 1 to N values as specified by the user. Initially Deferring process of transmission opportunities (contention slots) is checked and then as we need to treat the channel as busy for particular DATA packets based on circularity counter % circularity = 0 is checked. If the condition is true then that particular DATA packet refrained from the transmission by treating as channel busy else that DATA packet is send. After sending the DATA packet a further check is done to see if the DATA has been send successfully without collision. If no collision has taken place and ACK arrived for that particular node then transmission is successful, else the process again resumes from the beginning.

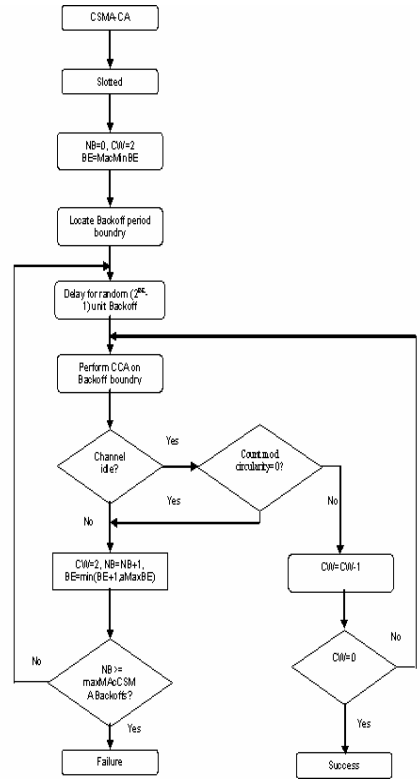


Figure 4: Flow chart for circularity in contention resolution

V. SIMULATION RESULTS

For simulations, the network simulator version 2.28 (NS-2) with the CMU wireless extension was used **[13]**. Also some appropriate changes are made to ~wpan/P802_15_4csmaca.cc. We used 250 Kbps for the channel bit rate. Packet size is 50 bytes unless otherwise specified. Each flow in the network transmits Poisson traffic. We performed the simulation with various network loads. For the radio propagation model, a free space path loss model is used. We do not consider fading in our simulations and all the simulation parameters for the star topology network with varying number of nodes (10 to 80 nodes) are listed as below. The circularity value has been assigned to the nodes for appropriate data packets. In the instance, the circularity values are varied and finally kept at optimum (k = 3), for which one gets maximum throughput as shown in the following graphs.

In Figure5 we observe that the numbers of collisions are increasing as the density of the nodes increases in the network. However the mechanism of circularity significantly reduces the number of data collisions as compared to the regular IEEE 802.15.4.

Super frame duration	960 symbols 15.36 ms
Rx-Tx / Tx-Rx turnaround time	12 symbols
Symbol duration	16 µs
Number of slots	16
Slot/symbol	60
Channel Used	Wireless
Antenna	Omni
Propagation	Free space
No. of Nodes	10-80
Traffic	Poisson
Packet Size	50 Bytes
No. of Channels	16
Routing Protocol	AODV
Channel Spacing	5Mhz
Symbol Rate	61.5
Data Rate	250Kb/s
Rx Max i/p Level	-20 dBm
Modulation	O-QPSK half-sine pulse
Simulation Period	200 s
O/P power (lowest Max.)	-3 db

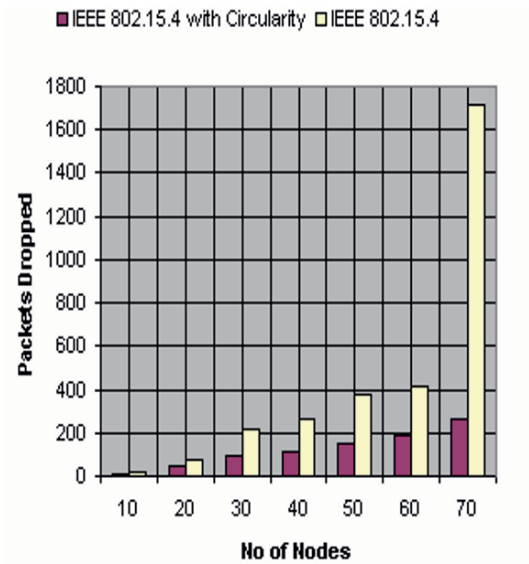


Figure 5: Number of collisions vs. Number of nodes

Owing to the paradigm of circularity, the Delay/Packet is considerably minimized in the varying scenarios of 10-80 nodes in comparison to that of normal IEEE 802.15.4 as shown in the Figure 6.

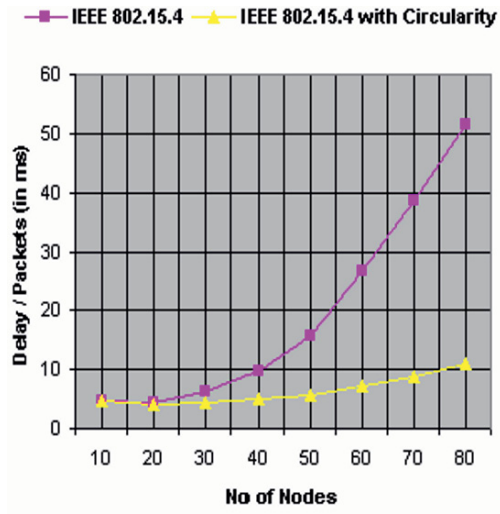


Figure 6: Delay/packet vs. Number of nodes

Finally the establishment of a notable increase in the throughput in terms of Packet/Second for the case of circularity is depicted in Figure 7.

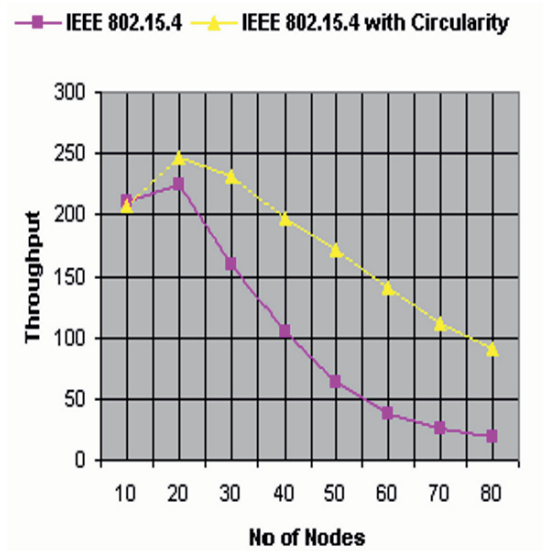


Figure 7: Throughput (PPS) vs. Number of nodes

Also from the figure one can observe that as the density of nodes increases beyond 70 or 80 plus nodes, the regular IEEE 802.15.4 fails to handle the bursty data packets and the throughput is almost reaching to zero.

This is caused due to enormously increased number of collisions. Whereas our new paradigm of circularity is still asymptotically maintaining some steady state value by reducing the number of data collisions and thus leads to the effective utilization of available Bandwidth.

VI. CONCLUSION

The formation of the mobile as well as sensor hosts in the hypothetical model of the LR-WPAN by combining the sensing element, the processing and transmitting device has been tacitly assumed. The inherent problems associated with the MAC sub layer are initially analyzed by cursorily estimating the time taken for single transaction. The consequent events caused by refraining transmission of DATA packets and the selective use of optimum value for circularity is visualized. These events yield better utilization of the available bandwidth by enhancing the contention resolution for the overall medium access to the hosts in the LR-WPAN. The results of the enhanced IEEE 802.15.4 MAC due to the mechanism of circularity are substantiated through NS-2 simulations for the assumed varying scenarios from ten to eighty nodes in LR-WPAN.

ACKNOWLEDGMENT

The author sincerely wishes to thank Prakash Satyanarayan for his notable encouragement and also places on record the valuable support provided by Tata Elxsi Limited.

REFERNCES

[1] E. Callaway, P. Gorday, Lance Hester, J.A. Gutierrrez, Marco Naeve, B. Heile and V. Bahl, “Home Networking with IEEE 802.15.4: A Developing Standard for Low-Rate Wireless Personal Area Networks”, IEEE Communications, pp. 70-77, Vol. 40, no 8, August 2002.

[2] J. A. Gutierrez, et al, “IEEE Standard for Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications for Low Rate Wireless Personal Area Networks (LR-WPAN) - Draft D16”, IEEE 2002.

[3] I. Howitt, “WLAN and WPAN Coexistence in UL Band,” *Transactions on Vehicular Technology*, vol. 50, pp. 1114-1124, 2001.

[4] Jelena Mi si c Vojislav and B. Mi si c Shairmina Shafi “Performance of IEEE 802.15.4 beacon enabled PAN with uplink transmissions in non-saturation mode – access delay for finite buffers”, Proceedings of the First International Conference on Broadband Networks, 2004.

[5] Nicholas F. Timmons and William G. Scanlon “Analysis of the Performance of IEEE 802.15.4 for Medical Sensor Body Area Networking”, IEEE SECON 2004.

[6] Jianliang Zheng and Myung J. Lee “Low Rate Wireless Personal Area Networks for Public Security”, Proc. VTC 2004, pp.4568-4572, vol. 6, September 26-29, 2004.

[7] IEEE P802.15-02/468r0 “Draft D17 Clause 7.6 Security Recommendation for Low-Rate IEEE 802.15.4 WPAN”, November 2002.

[8] Sofie Pollin, Mustafa Ergen, Sinem Coleri Ergen, Bruno Bougard, Liesbet Van der Perre, Francky Catthoor, Ingrid Moerman, Ahmad Bahai, and Pravin Varaiya “Performance Analysis of Slotted IEEE 802.15.4 Medium Access Layer”, IEEE Trans. Wireless Communications, 2005.

[9] Gang Lu, Bhaskar Krishnamachari, and Cauligi S. Raghavendra “Performance Evalution of the IEEE 802.15.4 MAC for Low-Rate Low-Power Wireless Networks”, Workshop on Energy-Efficient Wireless Communications and Networks (EWCN '04), held in conjunction with the IEEE International Performance Computing and Communications Conference (IPCCC), April 2004.

[10] Jianliang Zheng and Myung J. Lee “A Comprehensive Performance Study of IEEE 802.15.4”, IEEE Press Book 2004.

[11] R Bhakthavathsalam “Reinforcement of Privacy in 802.16 MAC Common Part Sublayer Using the Principle of Circularity”, Proc. CNIS06, pp.119-125, October 09-11, 2006, MIT, Cambridge, MA, USA.

[12] Mohammad Z. Ahmad, Damla Turgut, and R Bhakthavathsalam “Circularity-Based Medium Access Control in Mobile Ad-Hoc Networks”, ADHOC-NOW 2006, pp.337–348, LNCS 4104, Springer-Verlag Berlin Heidelberg.

[13] USC Information Sciences Institute, Marina del Rey, CA: Network Simulator – NS2. (http://www.isi.edu/nsnam/ns).

Wireless LAN Security Mechanisms at the Enterprise and Home Level

Bogdan Crainicu
"Petru Maior" University of Targu Mures
N. Iorga, no. 1
Targu Mures, MS 540088, ROMANIA
cbogdan@upm.ro

Abstract-**Securing Wireless LAN is a task in progress, being a constant concern for every network administrator. Choosing the right wireless LAN security mechanism is a real challenge, because it depends on what level of risk is allowable. Taking also into account the performance and interoperability issues, this paper covers the main wireless LAN security protocols from both the enterprise and home user perspective.**

I. Introduction

While the benefits of Wi-Fi are obviously, there are some security weaknesses that still are neglected. Unlike wired networks, the wireless packets can be easily captured and analyzed by unauthorized users. Once a wireless LAN (WLAN) has been located, whithout a strong security mechanism in place, a hacker may steal data or performs other malicious damage to WLAN.

But security measures and level of risk accepted by an organization are closely linked. An enterprise user will always have different security requirements than a home user. Security experts will always advice to switch from Wired Equivalent Privacy (WEP) [27] to IEEE 802.11i [29] regardless of the cost involved. WEP was the first solution designed to ensure a level of security as that of a wired network. WEP, as part of the IEEE 802.11 standard [27], aims to provide security services by protecting the confidentiality of user data. But in 2001, two papers [5], [22] demonstrated severe flaws in WEP, and, accordingly, additional measures have to be taken in order to protect data traffic.

IEEE 802.11i is an amendment to the IEEE 802.11 standard and supersedes the previous WEP security specification. In order to provide the ultimate robust security for the wireless networks, the 802.11i standard introduces new encryption key protocols including the Temporal Key Integrity Protocol (TKIP) [29] and the Advanced Encryption Standard (AES) [3].

This paper analyzes the protocols and mechanisms suitable for providing data confidentiality, (mutual) authentication, data integrity and availability in WLANs, as well as the security issues related to WLAN technologies, and focuses on the security measures that have to be implemented in accordance with the requirements of specific wireless users. As a guideline, the paper defines three WLAN security levels that carry out the users' needs from the home to the large enterprise:

- Level 1: Home WLAN Security
- Level 2: Small/Medium Network WLAN Security
- Level 3: Enterprise WLAN Security

II. The WEP Protocol And It's Vulnerabilities

The objective of WEP is to achieve data privacy to the level of a wired network. The IEEE 802.11 standard claims that WEP is "reasonably strong". In a relatively short time, the experts admit that WEP fails to meet its security goals, mainly because of WEP's use of 40-bit RC4 [20]. Because of its usage of Initialization Vector (IV), the WEP encapsulation is insecure whatever the key length, and the same remains true when any other stream cipher replaces RC4 [24]. The discovered flaws of WEP finally prove serious drawbacks of its design.

Reference [5] points out the risks of keystream reuse and demonstrates the vulnerability of CRC-32 checksum which fails to provide the required data integrity protection.

Reference [4] describes the flaws in two access control mechanisms and an eavesdropping attack against the IEEE 802.11 specified shared key authentication mechanism. The authors conclude that a comprehensive redesign of the standard is the best option.

Reference [12] presents several weaknesses in the key scheduling algorithm of RC4, describing a passive ciphertext-only attack. The authors identify a large number of weak keys, in which knowledge of a small number of key bits suffices to determine many state and output bits with nonnegligible probability. They also show how to reconstruct the secret key in WEP by analyzing enough WEP-encrypted packets.

Reference [22] describes an attacks against WEP based on [12], how to implement it and some optimizations to make attack more efficient. The attack used off of the shelf hardware and software and successfully recovered the secret key by observation of traffic.

Reference [23] brings forward a practical key recovery attack on WEP based on a partial key exposure vulnerability in the RC4 stream cipher presented in [12].

In summary, the security problems of WEP are:

- 24-bit Initialization Vectors (IVs) are too short (posibility of IV re-use, and consequently key re-use;

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 305–310.

with the birthday paradox, it is shown that there is a 50% chance of using the same IV value after only 4823 packets);
- the IEEE 802.11 standard specifies that changing the IV with each packet is optional;
- 32-bit CRC is linear (flipping bit n in the message results in a deterministic set of bits in the CRC that must be flipped to produce a correct checksum on the modified message) and does not prevent the modification of captured packets;
- inappropriate use of the RC4 cipher – WEP appends the IV to the secret key and [12] provides a solution which breaks the secret key after observing only a few million encrypted packets;
- manual key distribution is the facto standard for WEP;
- large key life-time;
- authentication is not mutual (one-way authentication only).

The proposed, quick fix to the WEP security problem was the Fast Packet Keying (FPK) solution from RSA, a mechanism designed to change the way IVs are derived. FPK uses a hashing technique that rapidly generates a unique pseudo-random RC4 key for each packet of data sent over the WLAN. The solution consists of an encryptor and decryptor that share a RC4 128-bit secret key called the temporal key (TK), an encryptor and decryptor that uses the RC4 stream cipher, and an IV value that is not used more than once with each TK.

Reference [25] proposes WEP*, a lightweight solution to the problem of access revocation from previously-authorized hosts, where key management is in the style of pay-TV systems: the access point (AP) periodically generates new keys, and these keys are transferred to the hosts at authentication time. Moreover, the WEP* solution requires no additional equipment beyond the AP and hosts.

The IEEE has proposed an updated standard, WEP2, to address the WEP shortcomings. WEP2 uses 128-bit encryption and a 128-bit IV. But this stopgap measure is still based on the same RC4 cipher. Moreover, the approach of static keys remain, the cheksum system is the same and then it is not immune to the Fluhrer-Mantin-Shamir (FMS) attack. Because the RC4 keystream is longer, it was expected that the difficulty in capturing the WEP key will grow in a linear fashion, not exponentially. Even the attacker spends a few more hours to collect the necessary packets, the attack is very viable.

III. WPA AND WPA2/IEEE 802.11I

The Wi-Fi Protocted Access (WPA) standard was created by the Wi-Fi Alliance in response to serious security weaknesses discovered in the WEP protocol, providing more secure encryption and data integrity while the IEEE 802.11i standard was being finished. WPA resolves the issue of weak WEP IV, and provides a method of guaranteeing the integrity of the messages using the Temporal Key Integrity Protocol (TKIP) to enhance wireless security. In fact, TKIP includes three elements that adapt the WEP to address the known drawbacks: a Message Integrity Code (MIC) called Michael [11] to defeat forgeries, a packet sequencing discipline to defeat replay attacks, and a per-packet key mixing function to prevent FMS attack [15]. The two-phases per-packet key is created by mixing together a combination of elements, including the Pairwise Transient Key (PTK), the MAC address of the transmitting node and the 48-bit TKIP IV. Additionally, the TKIP IV is used as a sequence counter which is incremented sequentially. This solves the collision attacks and mitigates the replay attacks.

One of the strongest part of WPA is the capability to provide strong authentication via the IEEE 802.1X standard [28]. This port-based network access control standard makes use of the physical access characteristics of 802 LAN in order to offer a means of authenticating and authorizing devices attached to a LAN port that has point-to-point connection characteristics [28]. On WLAN, IEEE 802.1X provides also a method of dynamic key distribution by the generation of two key sets: pairwise keys and groupwise keys. With 802.1X, the user (supplicant) is not authenticating to the wireless access point (authenticator), but to the LAN, through an authentication server behind the access point such as RADIUS [19] or DIAMETER [6], that integrates some form of user database (native RADIUS, Active Directory, LDAP).

802.1X is very flexible because it is base on Layer 2 Extensible Authentication Protocol (EAP) [2], which makes possible the several types of authentication credentials. 802.1X encompasses a range of EAP authentication methods, such as: MD5 (Message Digest 5), TLS (Transport Layer Security), TTLS (Tunelled TLS), LEAP (Lightweight EAP), PEAP (Protected EAP), EAP-PSK (Pre-Shared Key), EAP-IKEv2 (Internet Key Exchange Protocol version 2), EAP-FAST (Flexible Authentication via Secure Tunneling), EAP-SecurID (Secure ID hardware token), EAP-SIM (GSM Subscrieber Identity), EAP-AKA (UMTS Authentication and Key Agreement). Through the use of EAP, support for a number of authentication schemes may be added: Kerberos, Smart Cards, Public Key Encryption, One Time Passwords.

WPA is in fact a subset of IEEE 802.11i security specifications, based on draft 3 of the IEEE 802.11i amendment. A very important aspect of WPA is that it was designed to be a replacement for WEP protocol without requiring hardware upgrade. Unfortunately, legacy WEP hardware cannot understand 48-bit TKIP IV. To solve this problem, the 48-bit IV is divided into a 16-bit part padded to 24 bits to produce a 24-bit IV, and a 32-bit part used in the TKIP per-packet key mixing. Therefore, TKIP is backward compatible with WEP and does not require hardware replacements.

TKIP has its weaknesses as well. [26] shows that the security of the MIC completely breaks down if a single message with its MIC value is exposed, and how a related message attack can can expose the MIC value if a TKIP implementation ever reuses IV values. [16] analyzes some

weaknesses in WPA authenticator key management state machine and proposes the countermeasures to overcome these problems. [17] describes another weaknesses in the key scheduling in WPA. This paper proves that the whole security in WPA relies on the secrecy of all packet keys - given one packet key it is possible to find the MIC key and given two packet keys with the same 32-bit IV an attacker can do anything the legitimate user can, for the duration of the temporal key.

The next protocol, WPA2 (the Wi-Fi Alliance refers to their fully ratified, interoperable 802.11i implementation as WPA2), based on Robust Security Network Architecture (RSNA) concept, implements the mandatory elements of IEEE 802.11i. In particular, in addition to TKIP and the Michael algorithm, WPA2 introduces a new AES-based algorithm called Counter Mode with Cipher Block Chaining Message Authentication Code Protocol (CCMP), which combines the counter mode for data confidentiality and the Cipher Block Chaining Message Authentication Code (CBC-MAC) for data integrity [14], [29].

AES Counter Mode is a block cipher that encrypts 128-bit blocks of data at a time with a 128-bit encryption key. The CBC-MAC algorithm produces a MIC that provides data origin authentication and data integrity for the packet. CCMP uses a Packet Number field included in the WPA2-protected wireless frame as a counter and also incorporated into the encryption and MIC calculations to provide replay protection. Like WPA, WPA2 relies on IEEE 802.1X for strong authentication and distribution of security keys, requiring the calculation of Pairwise Transient Keys (PTKs) through a 4-way handshake.

Both WPA and WPA2 can operate in either of PSK or Enterprise mode:

- WPA-PSK: RC4-TKIP and static key management (PSK);
- WPA-Enterprise: RC4-TKIP and automatic key management (IEEE 802.1X and EAP);
- WPA2-PSK: RC4-TKIP, AES-CCMP and static key management (PSK);
- WPA2-Enterprise: RC4-TKIP, AES-CCMP and automatic key management (IEEE 802.1X and EAP).

PSK mode is especially suited for home networks, where a strong (minimum 20 characters) pre-shared key can be shared amongst wireless clients and access point. Enterprise mode is more suitable for large organizations using an external authentication method. Both classes of WPA/WPA2 require encrypted authentication to occur prior to allowing encrypted data to be sent over the radio.

IV. Levels Of WLAN Security

After analyzing the characteristics of the most important WLAN security standards, protocols and mechanisms, we will define three levels of WLAN security that can assist in deployments of a strong WLAN security policy.

A. Level1: Home WLAN Security

Security must be an important concern on any WLAN. Generally, the home users are still using devices and software that support only WEP. For low volume of traffic, WEP encryption should be enough, because an attacker needs a few millions of packets to break a WEP key. But with the actual computational power and the latest WEP cracking tools, the WEP key could be recovered in minutes. An immediate improvement is to call the dynamic refresh of keys. Rather than a single static WEP keys for all packets on the wireless network, dynamic WEP solutions utilizes different keys.

Neverthless, the entirely wireless platform should be upgraded to WPA, at least. If the devices' firmware or drivers cannot be upgraded from WEP to WPA, they must be taken out. Today, WPA/WPA2 certified access points are no more expensive, and WPA/WPA2 support in personal operating systems such as Windows XP, Linux and MacOS is widely available.

What about the authentication options at this level? It is very unlikely that a Public Key Infrastructure (PKI) and/or IEEE 802.1X take here place. The solution is WPA-PSK mode in the form of a single shared secret (password) which is included into both access point and into each wireless devices. The use of WPA-PSK is recommended for home use only, since the pre-shared key is used as the Pairwise Master Key (PMK) impersonation between stations or a station impersonating an AP is possible. Although sometimes is annoying and time consuming, a proper security practice says to change frequently the passwords on every wireless device. And moreover, because WPA-PSK is vulnerable to offline dictionary attack, a strong random passphrase is a must – at least 20 digit string of very random characters.

A salutary project called TinyPEAP [30], available as yet for beta testing only, implements a very small RADIUS server that supports PEAP/MSCHAPv2 authentication. TinyPEAP provides the benefits of 802.1X and PEAP security without the complexity of deploying a dedicated RADIUS server. TinyPEAP runs the RADIUS server and user database directly on the access point itself, eliminating the need for an external server, and utilizes PEAP and MSCHAPv2 for user authentication. Its ability to run on hardware found in cheap wireless access points gives it a significant advantage when price is a concern, making it an attractive authentication option for small WLAN implementations.

WPA-PSK is not a long term security solution because of the RC4 algorithm and static key management. Parts of WPA are also weak, being possible to find the MIC keys and the Temporal Key (TK). One of the advantages of WPA is that it is backward compatible with old WEP and does not require wireless hardware upgrades. Instead, the AES algorithm of WPA2 needs a separate chip hardware implementation and therefore the upgrade to WPA2 typically requires a hardware upgrade or replacement.

B. Level 2: Small/Medium Network WLAN Security

The starting point of Level 2 goes one step ahead of Level 1 by adding strong authentication to the WLAN security toolbox. The WEP implementations must be excluded at this level. The IEEE 802.1X standard addresses authentication issues, providing a method to protect the network behind the access point based on Extensible Authentication Protocol (EAP). The challenging part in deploying 802.1X on a wireless network is to decide on the type of authentication that will be implemented, because the final decision will guide the product choices.

From among many forms of EAP, three EAP methods could be considered appropriate for this level, where PKI are not deployed: EAP-PSK, EAP-TTLS and PEAP. EAP-PSK is still an Internet draft and relies on a shared secret, offering a litle more security than EAP-MD5. After exchanging four messages in the authentication process between a client and a server, and if the authentication procedure ends successfullly, EAP-PSK finally provides a protected communication channel. Too little open-source implementations of EAP-PSK are offered as yet both for the client side (under Windows XP and Linux) and for the server side (under Linux using FreeRadius).

EAP-TTLS extends EAP-TLS and offers very good security. A big advantage of EAP-TTLS is that it it widely deployed across platforms, while not requiring that each user be issued a certificate. Thus, the client does not need be authenticated via a signed certificate to the server, but only the server to the client.

PEAP is similar in design to EAP-TTLS in that it does not require a certificate on the client side. EAP-PEAP is less flexible than EAP-TTLS and it does not support the range of inside-the-tunnel authentication methods that EAP-TTLS supports. On the other hand, PEAP's major advantage is support from Microsoft, and therefore, built-in support from the operating system [13].

IEEE 802.1X divides the network into three well-defined entities: the supplicant, which wants to join the network, the authenticator, which controls access, and the authentication server, which takes the authorization decisions. The authentication server is usually a RADIUS or DIAMETER server, which has to be placed on an isolated network segment that only the RADIUS/DIAMETER clients can reach it. Thus, to deploy EAP-TTLS or EAP-PEAP, the wireless network needs a RADIUS or DIAMETER authentication server [1], [8], [10]. The DIAMETER protocol is derived from the RADIUS proocol with some improvements, and is generally credited as the next generation Authentication, Authorization and Accounting (AAA) protocol.

Access points must implement a RADIUS client which sends connection parameter informations and user credentials to a RADIUS server. The RADIUS server authenticates and authorizes the RADIUS client request, and sends back a RADIUS message response. RADIUS messages are never sent between the supplicant and the access points. The access points share a secret with the RADIUS server which makes them trusted for it; therefore, a supplicant which trusts a RADIUS, also trusts access points.

Although RADIUS was not initially intended to be a wireless security authentication method, it is the default authentication protocol used by the IEEE 802.1X. For EAP-TTLS and EAP-PEAP, the RADIUS server must implement a server side digital certificate. For small/medium WLAN infrastructure, it could be expensive to buy digital certificates from a third party Certification Authority (CA) or even to install a PKI platform which integrates an internal CA server responsible for issuing local certificates. The help comes in form of self-signed digital certificates – an elegant way avoiding the use of CAs by organizations that cannot afford a dedicated PKI CA infrastructure.

If all security components are in place and functional, dangers for this security level remain, such as stealing user credentials. But security is all about trust, and the social engineering attacks exploit, at every level, the human vulnerabilities.

Provisions for additional safeguards could be introduced at IP level, such as IPsec, but every new security measure will increase the complexity of the wireless infrastructure and most probably the resource usage and the cost will increase as well.

C. Level 3: Enterprise WLAN Security

At Level 3, the WLAN security functions have to be chosen according to the key features needed by an enterprise organization: client mobility, authentication services, expandability, firewall mechanism, VPN deployment, QoS support, intrusion detection/prevention system. There is about a fully-deployed RSN design, proof against attacks such as malicious association, MAC spoofing, man-in-the-middle, session hijacking, and denial-of-service.

The implementation is absolutely based on WPA2-Enterprise specifications: 802.1X along with EAP-TLS or PEAP-EAP-TLS (PEAP-EAP-TLS is very similar in operation to the original EAP-TLS, but provides slightly more protection due to the fact that portions of the client certificate that are unencrypted in EAP-TLS are encrypted in PEAP-EAP-TLS) and RADIUS/DIAMETER authentication server, automatic key management and AES-CCMP encryption protocol. Although the CCMP provides much stronger security services than TKIP, it requires additional hardware to improve encryption performance [7].

One of the most challenging aspect is the PKI deployment process. With EAP-TLS and PEAP-EAP-TLS, both the client and the server must be allocated a digital certificate signed by a qualified CA that they both trust, and not self-signed certificates. For a medium enterprise organization which utilizes certificates that only need to be trusted internally, it is recommendable a two-tier PKI hierarchy with an offline root CA and an online issuing CA. For a large eneterprise organization, a three-tier PKI approach, with an offline root CA (third-party root or internal root), an offline subordinate/intermediate CA and an online issuing CA, reaches a high level of redundancy: if a

subordinate/intermediate CA is compromised, only the certificates it has signed need to be replaced. An subordinate/intermediate CA's security requirements are the same as for the root CA because an intermediate CA provides CA certificates to online issuing CA.

Some final considerations about root CA: the root CA role is very important because it is trusted by all organization entities; if the root CA is ever compromised or its private key and certificate are changed, the whole PKI deployment must rebuild, and even for an enterprise organization the costs could be very high. For this reason, it is recommendable to protect the root CA private key as much as possible.

Measuring, tracking and reporting security performance is a critical task. PKI infrastructure management, along with security policy, audit management, incident response, business continuity, risk assessment and user awarness training are parts of the security strategic plan of every enterprise organization.

This level of security can be further improved. In many cases, the digital certificates (private keys) are stored on network-connected computers. Viruses, worms, malicious programs are capable of stealing private keys and other security credentials, and even an unforeseeable computer stealing could have very long-term security consequences. This is way the root CA machines and the subordinate/intermediate CA machines are physically secured and almost never connected to the network. In order to defense these threats, the use of Hardware Security Modules (HSMs) are very advisable from the CA point of view. The task of the HSM is to securely generate and store long term secrets (private keys), and to physically protect the access to and use of those secrets over time. Moreover, most HSMs provide hardware protection of symmetric keys and harware cryptographic accelerator capabilities.

On the client side, PKI tokens and PKI smart cards are devices that contain the users digital certificates and private keys protected by a PIN code. The Trusted Platform Module (TPM)-embedded computers offer facilities for secure generation of cryptographic keys, protected storage of keys through hardware, hardware protection of authentication data, protection of authentication credentials by binding them to platform, the ability to limit the use of keys, PKI-based methods for platform authentication, as well as a hardware random number generator.

The biometric PKI solutions provide the strongest authentication mechanism possible, being the culmination of development of secure cryptographic key generation and storage with biometric access control: fingerprint biometric technology, certificates embedded with DNA personal ID, etc. These military-like security systems are expensive and very hard to compromise. And probably, the previously-mentioned social engineering aspects remain the only attack aproach.

As a strict rule, a certificate must be revoked immediately when: it has been compromised/lost/stolen or the owner of the certificate no longer has the right to use the certificate.

V. Conclusions

WLANs provide concrete benefits over wired networks, including mobility, simplicity, flexibility, scalability and not at least the installation and maintenance costs. But these obvious advantages incur serious risks that comes from well-known aspects such as authentication, access control, confidentiality, integrity, and that have to be very carefully addressed [21]. Many options are available to protect WLANs and a good understanding of each of these options is desirable in order to identify the security vulnerabilities and mitigate the threats.

The paper puts forward a WLAN security guide based on the actual technologies and protocols, the discovered WLAN security drawbacks and weaknesses, the size of the WLAN and the WLAN user requirements. Taking into account that creating a secure wireless network is an ongoing process, the paper defines three WLAN security levels that follow the WLAN security best practices. From the home user to the enterprise user, the objectives remain the same: the benefits must outweigh the risks in a predominant manner.

References

[1] Aboba B., Calhoun P., "RADIUS (Remote Authentication Dial In User Service) Support For Extensible Authentication Protocol (EAP)", Request for Comments: 3579, Network Working Group, 2003

[2] Aboba B., Blunk L., Vollbrecht J., Carlson J., Levkowetz H., "Extensible Authentication Protocol (EAP)", Request for Comments: 3748, Network Working Group, 2004

[3] *Advanced Encryption Standard*, NIST FIPS PUB 197, November 2001

[4] Arbaugh, W., A., Shankar, N., Justin Wan, Y., C., "Your 802.11 Wireless Network has No Clothes", in *Proc. IEEE Int. Conf. Wireless LANs and Home Networks*, 2001, pp. 131-141

[5] Borisov, N., Goldberg, I., Wagner, D., "Intercepting Mobile Communications: The Insecurity of 802.11", in *Proc. 7th ACM Conf. Mobile Computing and Networking*, Rome, Italy, 2001, pp 180-189

[6] Calhoun P., Loughney J., Guttman E., Zorn G., Arko J., "Diameter Base Protocol", Request for Comments: 3588, Network Working Group, 2003

[7] Chen, J., C., Wang, Y., P., "Extensible Authentication Protocol (EAP) and IEEE 802.1x: Tutorial and Empirical Experience", *IEEE Communication Magazine*, Vol. 43, No. 12, December 2005, pp supl.26-supl.32

[8] Congdon, P., Aboba, B., Smith, A., Zorn, Roese, J., "IEEE 802.1X Remote Authentication Dial In User Service (RADIUS) Usage Guidelines", Request for Comments: 3580, Network Working Group, 2003

[9] Edney, J., Arbaugh, W., A., "Real 802.11 Security: Wi-Fi Protected Access and 802.11i", Addison Wesley, 2003

[10] Eronen P., Hiller T., Zorn G., "Diameter Extensible Authentication Protocol (EAP) Application", Request for Comments: 4072, Network Working Group, 2005

[11] Ferguson, N., "Michael: An Improved MIC for 802.11 WEP", Available: http://grouper.ieee.org/groups/802/11/Documents/DocumentHolder/2-020.zip

[12] Fluhrer, S., Mantin, I., Shamir, A., "Weaknesses in the Key Scheduling Algorithm of RC4", in *Proc. 8th Workshop Selected Areas in Cryptography*, LNCS 2259, 2001, pp 1-24

[13] Gast, M., "802.11 Wireless Networks: The Definitive Guide", Second Edition, O'Reilly, 2005

[14] He, C., Mitchell, J., C., "Security Analysis and Improvements for IEEE 802.11i", *The 12th Annual Network and Distributed System Security Symposium (NDSS'05)*, February 2005, pp 90-110

[15] Huang, J., Seberry, J., Susilo, W., Bunder, M., "Security Analysis of Michael: The IEEE 802.11i Message Integrity Code", *Lectures Notes in Computer Science*, Vol. 3823, Springer Berlin/Heidelberg, November 2005, pp 423-432

[16] Kang, Y., S., Oh, K., Chung, B., Chung, K., Nyang, D., "Analysis and Countermeasure on Vulnerability of WPA Key Exchange Mechanism", *Lectures Notes in Computer Science*, Vol. 3090, Springer Berlin/Heidelberg, August 2004, pp 915-924

[17] Moen, V., Raddum, H., Hole, K., J., "Weaknesses in the Temporal Key Hash of WPA", *Mobile Computing and Communications Review*, Vol. 8, No. 2, April 2004, Papers from MC^{2}R Open Call, pp 76-83

[18] Ou, G., "Wireless LAN security guide, Security for any organization large or small", January 2005, Available: http://www.lanarchitect.net/ Articles/ Wireless/SecurityRating/

[19] Rigney C., Willens S., Rubens A., Simpson W., "Remote Authentication Dial In User Service (RADIUS)", Request for Comments: 2865, Network Working Group, 2000

[20] Rivest, R., L., "The RC4 Encryption Algorithm", *RSA Data Security, Inc*., 1992. (Proprietary).

[21] Shunman W., Tao, R., Wang, Y., Zang, J., "WLAN and it's security problems", in *Proceedings of the Fourth International Conference on Parallel and Distributed Computing, Applications and Technologies*, 2003, pp 241-244

[22] Stubblefield, A., Ioannidis, J., Rubin, A., D., "Using the Fluhrer, Mantin, and Shamir Attack to Break WEP", *AT&T Labs Techical Report TD-4ZCPZZ*, August 2001

[23] Stubblefield, A., Ioannidis, J., Rubin, A., D., "A Key Recovery Attack on the 802.11b Wired Wquivalent Privacy Proocol (WEP)", *ACM Transactions on Information and System Security*, Vol. 7, No. 2, May 2004, pp 319-332

[24] Walker, J., "Unsafe at any key size: an analysis of the WEP encapsulation," *Tech. Rep. 03628E, IEEE 802.11 committee*, March 2000, Available: http://grouper.ieee.org/groups/802/11/Documents/ DocumentHolder/0-362.zip.

[25] Wool, A., "Lightweight key management for IEEE 802.11Wireless LAN's with key refresh and host revocation", *IEEE 802.11 TGi working group*, Available: http://grouper.ieee.org/groups/802/11/Documents/ DocumentHolder/2-411.zip

[26] Wool, A., "A Note on the Fragility of the "Michael" Message Integrity Code", *IEEE Transactions on Wireless Communications*, Vol. 3, No. 5, September 2004, pp 1459-1462

[27] *IEEE Standard for Information Technology-Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, ANSI/IEEE Std 802.11, 1999 Edition (R2003)

[28] *IEEE Standard for Local and metropolitan area networks - Port-Based Network Access Control*, IEEE Std 802.1X-2004

[29] *IEEE Standard for Information technology- Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 6: Medium Access Control (MAC) Security Enhancements*, IEEE Std 802.11i-2004

[30] *TinyPEAP Project*, Available: http://www.tinypeap.org/

Synchronization Solution for the TDSC-UWB Detection Method

Charbel Saber, Roger Lamberti, Claude Gimenes
GET-Institut National des Télécommunications
9, rue Charles Fourier, 91011 Evry France

Abstract— **This paper proposes a simple synchronization scheme for a new TR-UWB detection method called Time Delayed Sampling & Correlation (TDSC). The TDSC concept is based on inter-correlation between two waveforms captured by two time delayed samplers. The objective is to determine a synchronization scheme with lowest complexity for silicon integration without affecting the BER performance. A multiple correlator acquisition structure in conjunction with a training sequence and a smoothing procedure is used to estimate the signal correlation. Acquisition speed and BER performance are analyzed using IEEE UWB channel models. Results show that good performance are obtained using few correlators and fast acquisition.**

Ultra-wideband (UWB), Impulse radio (IR), Transmitted reference (TR), Time Delayed Sampling & Correlation (TDSC), Synchronization.

1 INTRODUCTION

Ultra-wideband (UWB) impulse radio is a carrierless system based on transmission and reception of very short duration pulses. Due to the large bandwidth of UWB signals, the indoor channel is extremely frequency selective, resulting in hundreds of resolvable multipath components [1]. This resolution provides diversity that can be used to enhance performance of the receiver. The challenge at the receiver is how to efficiently capture the energy from all these multipaths.

The transmitted pulses may be distorted by the channel. A coherent correlation receiver [2] demodulate the received signal by correlating it with a well designed template signal. So a receiver filter matched to the transmitted pulse in coherent detection may not work well if the pulse shape is distorted by the channel. The correlation receiver also requires precise template signal design, which is difficult due to the distortions of the signal during transmission [3].

To correlate the received signal with the template, the receiver needs to achieve a synchronization with inaccuracy much smaller than one pulse duration. A Rake receiver has problems in channel estimation and synchronization and requires to be implemented a large number of fingers to take full advantage of the available signal energy [2], which is prohibited in practice due to high complexity and high cost.

A Transmitted Reference (TR) technique [4] does not require explicit channel estimation or a large number of fingers, and is robust to possible channel distortion on pulse shape. In UWB TR system, a reference pulse and a modulated data pulse separated by a delay of T_D seconds constitute a doublet to represent one bit information. Normally the delay T_D is larger than the maximum delay spread of the channel plus one pulse duration to avoid interference between the received pulses of the same doublet. For UWB channels [5] with large dispersion, the delay needed between reference and data pulses requires long delay lines which can not be integrated on chip [6]. Add that its difficulties of realization present several problems, especially on delay precision and stability [7]; which hence, affect the detection performance, small synchronization error leads to great performance degradations.

To overcome this problem a new idea [8] for TR transmission called TDSC has been recently introduced by the authors [9]. The concept of TDSC is based on two waveform samplers and a digital delay which outperforms the traditional TR receiver.

In the TR system symbol synchronization must be done to find the maximum of the correlation. Several studies have been done about the timing synchronization in different configurations of receivers [10][11][12]. In this paper we present a simple and efficient structure for the timing synchronization with the TDSC receiver. Its main advantage is that a good synchronization is achieved with very few symbols.

Our final objective is to develop a low complexity, low power and low cost system with the TDSC principle. The energy consumption will be improved by reducing the complexity of the synchronization stage, with taking into account the global performance.

We first describe the architecture of TDSC in section 2, and the system model in section 3.

In section 4, the synchronization architecture specific to the TDSC structure is presented with a synchronization process for information packets transmission.

Section 5 studies the BER performance and the synchronisation probability with a smoothing procedure to improve the signal-to-noise ratio.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 311–316.

According to BER performance, simulation shows in section 6 that the TDSC receiver demonstrates good robustness to the low accuracy of the synchronization procedure.

Finally, conclusions are drawn in section 7.

2 TDSC - A NEW APPROACH FOR TR UWB SIGNAL DETECTION

The TDSC principle [8] is a different approach of a TR receiver based on digital commands and analog samplers.

The overall structure of TDSC is shown in Figure 1. The TR-UWB signal is sampled by two analog waveform samplers (sampler A & B) with pre-defined delayed digital commands which matches the exact time delay T_D between the pulses in a TR-UWB frame, as shown in Figure 2. The signal detection will be done after a correlation time T_W between the two TR-UWB signal samples.

The analog waveform sampler [13] uses an asynchronous delay line composed of simple inverters to generate the sampling commands at the different moments. This asynchronous implementation permits a sampling rate much higher with much lower power consumption than that of a synchronous design by using a global and explicit clock. The role of the broadband delay line is replaced by a time delayed double sampling which removes completely the need of an explicit analog signal delay line and all the design difficulties associated with [6] and [7]. In this case, the delay is only applied to the sampler's command digital signals, which can be generated easily and programmable in real-time from the system clock with an extremely high precision and high stability. So the parameter T_D in TR-UWB can be used for channel coding and multiplexing.

The correlation is done only when the two samplers are filled, and the output of the correlator correspond to a sample of the received signal correlation. As this sample is used for signal detection the system needs a timing synchronization to get the maximum of the signal energy.

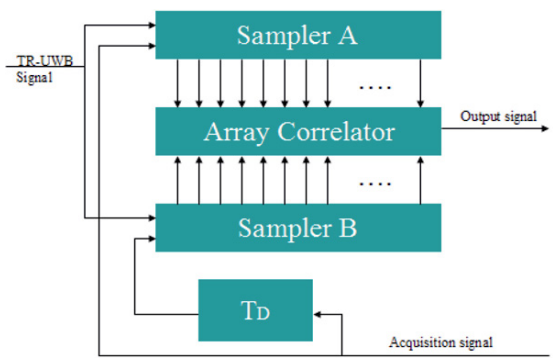


Figure 1. Structure of TDSC receiver, based on two samplers temporally delayed.

3 SIGNAL MODEL

The UWB signal used for TDSC follows the general TR signal transmission. We suppose here, for simplicity, a single user scenario with no frame protocol. An emitted bit or symbol is a doublet $d(t)$ in BPSK modulation given by :

$$d(t) = g(t) \pm g(t - T_D) \tag{1}$$

where $g(t)$ is a transmitted pulse that is nonzero value for $t \in [0, T_P]$. The energy of the pulse is Eg, the half of the doublet energy.

The first pulse is the reference and T_D later the second is a binary antipodal data modulated waveform.

The design of the pulse shape takes into account the FCC spectrum mask and is compatible with the sampling rate at the TDSC receiver. We use a simple Gaussian Hanning windowed pulse with standard deviation $\sigma = 1\text{ns}$ and $T_p = 8\text{ns}$.

A received symbol after the antenna and the amplifier is :

$$d_r(t) = g_r(t) \pm g_r(t - T_D) + n(t) \tag{2}$$

Where $g_r(t)$ is the received pulse through the UWB multipath channel proposed by IEEE 802.15.3a [14], see Table I, and $n(t)$ an AWGN limited to the signal band by a filter.

$$g_r(t) = g(t) * h_c(t) \tag{3}$$

$$h_c(t) = X_c \sum_{l=1}^{L} \sum_{k=1}^{K} \alpha_{kl} \delta(t - \tau_{kl} - T_l) = X_c \sum_{i=1}^{Np} \alpha_i \delta(t - \tau_i) \tag{4}$$

Where α_i and τ_i are the amplitude and delay of the *i*-th path, X_c represents the log-normal shadowing, and subscript *c* refers to the channel realization.

TABLE I. IEEE 802.15.3A INDOOR CHANNEL CHARACTERISTICS

Channel models	CM1 LOS	CM2 NLOS	CM3 NLOS	CM4 NLOS
RMS delay (ns)	5.5	8.5	15	25.5

The received signal is then : $$r(t) = \sum_{k=-\infty}^{+\infty} d_r(t - kT_s) \tag{5}$$

At rate T_e, the TDSC (Fig.1) first samples the reference waveform in the sampler A and after the delay T_D samples the data waveform in the sampler B, and finally at the instant τ, provides the decision variable $c(\tau)$ at the correlator output :

$$c(\tau) = \sum_{i=1}^{N_W} A_i B_i \underset{-1}{\overset{1}{\gtrless}} 0 \tag{6}$$

Where A_i and B_i are the samples of sampler A and B respectively, N_W is the number of taps of the correlator.

The integration time of the correlator is :

$$T_W = N_W T_e \tag{7}$$

The time τ is the delay between the beginnings of the received symbol at time $t = kT_s$ and the sample at the output of the correlator. With the TDSC, the pulse synchronization and the channel estimation are fulfilled by the doublet, the symbol detection only needs symbol synchronization. This synchronization process requires to estimate the time shift and to find the maximum of the correlation curve $c(t)$ which is equivalent to the ML estimate with the window N_W .

4 SYNCHRONIZATION

We propose here a structure of the TDSC correlator receiver for fast synchronization, it uses a training sequence and could be adapted in blind mode. The structure and the procedure are presented for one user in BPSK detection and they could be extended to various PPM and multi users configuration.

4.1 Synchronization architecture

We suppose that information is structured into packets by the MAC module. The algorithm is based here on the presence of a fixed synchronization preamble at the beginning of each packet, and assumes that synchronization is maintained for the rest of the packet. The packet length must be shorter than the coherence time of the shorter channel. To increase the part of the data, a blind detection of the channel coherence can be added to the algorithm.

The receiver architecture is presented in Figure 3, the receiver utilizes *N* identical correlators spaced over wide TDSC waveform samplers. Each correlator is shifted from the precedent by the time delta = Δ , it is preferable to use a regular sampling rate i.e. equal spacing Δ between the correlators. These correlators produce N samples of the correlation $c(t)$ (Fig.2).

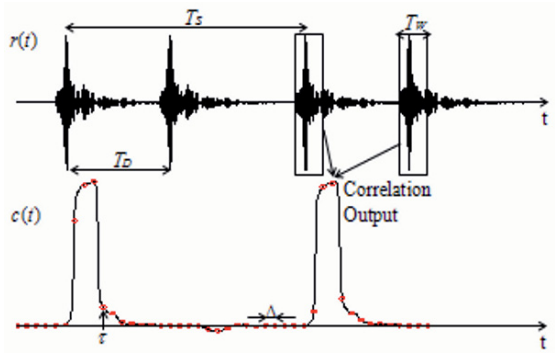


Figure 2. Free Noise Received Signal and Correlation

The interval $\Delta_{1,N}$ correspond to the sampler temporal spread T_{spl} minus the integration time T_W of a correlator :

$$\Delta_{1,N} = (N-1)\cdot\Delta = T_{spl} - T_W \quad (8)$$

These *N* samples cover the time $N\cdot\Delta$ wich is only a part of the whole symbol duration T_s ; this means that the entire correlation must be estimated with at least K successive acquisitions:

$$K = T_s / N\Delta \quad (9)$$

A continuous acquisition is not possible due to the TDSC principle i.e. the registers are filled and then correlated with a minimal periodicity of $T_D + T_{spl}$ where T_{spl} is inversely proportional to the frequency bandwidth of the system. For example in 350nm CMOS technology we can have $T_{spl} = 31$ns for a 6GHz bandwidth and $T_W = 15$ns (see [15]). To avoid ISI the symbol period T_s must be greater than the delay spread of the channel plus the time delay T_D between the two samplers, furthermore to avoid inter pulse interference (IPI) the delay T_D must be also greater than the delay spread. So a comfortable spacing leads to $T_s = 300$ns for a CM4. This means that with $N = 5$ correlators and $T_D = 150$ns the entire correlation is estimated within $K = 15$ acquisition symbols, this is very a short time raw estimation which will allow long smoothing for noise reduction. The number of symbols can be dramatically reduced with shorter delay and if the technology allows multiple acquisitions during the symbol time T_s, for example with $N = 5$, $T_D = 30$ns, $T_s = 180$ns and 3 acquisitions by symbol gives $K = 9$ and only 3 symbols.

4.2 Synchronization process

To estimate the position of the correlation maximum we have to sample this correlation over a symbol interval T_s. Assuming the case where only one acquisition is available during the symbol time, we can reconstitute the entire correlation by delaying the sampling command of $N\Delta$ seconds at each symbol. The successive delays are calculated by the DSP and achieved by the digital delay synchro in Figure 3. The complete correlation is obtained after a cycle of $K = T_s / N\Delta$ symbols. At the *k*th acquisition, the sample output of the *n*th correlator is given by :

$$c(\tau_m) = c_{nk} = \sum_{i=1}^{N_w} A_{ink} B_{ink} \quad (10)$$

with : $n = 1, N;\ k = 1, K$ and $m = n + (k-1)N$ (11)

τ_m is the *m*th sampling point during the symbol interval. A_{ink} and B_{ink} are the signal samples in the samplers A and B at the respective times t_{ink} and $t_{ink} + T_D$ where :

$$t_{ink} = t_0 + iT_e + (n-1)\Delta + (k-1)(N\Delta + T_s) \quad (12)$$

t_0 is the arbitrary time position of the 1st acquisition command.

The data are stored in a matrix $\mathbf{C}_{N\times K}$, at each acquisition the parallel outputs of the N correlator are stored in one column of the matrix, the succeeding columns are filled at each symbol with the time shift of $N\Delta$. In the case of multiple acquisitions in one symbol the matrix is filled by blocs. After

the K acquisitions the matrix is complete and the process can be iterated for smoothing.

The greatest term c_{nk} of the matrix $\mathbf{C}$ gives both the position estimate and the value of the correlation maximum.

$$\hat{\tau}_{\max} = \arg\max_{n,k}(c(\tau_{nk})) \quad \text{and} \quad \hat{c}_{\max} = \max_{n,k}(c_{nk}) \qquad (13)$$

The delay of the maximum relative to the received signal is then :

$$\hat{\tau}_{\max} = \tau_1 + (\hat{n}-1)\Delta + (\hat{k}-1)N\Delta \qquad (14)$$

After estimation, at the end of the synchronization preamble, a corresponding delay in the acquisition command is applied, this bring the maximum into the first column of the matrix. At this time, during the detection phase, only one acquisition by symbol is needed which uses the estimated maximum correlator output. Moreover, at the same time as the detection, a blind procedure could be applied to track little changes in the channel and maintain the smoothed maximum in this column; this can also be used to detect a more important change in the channel.

4.3 *Precision*

The accuracy of the synchronization time is related to the amount of Δ, or the number of correlators for a specific sampler length. The compromise is to determine the less number of correlators for low complexity integration and to keep a good detection performance.

Without noise the precision on the optimal synchronization time is

$$\hat{\tau}_{\max} - \Delta < \tau_{\max} < \hat{\tau}_{\max} + \Delta \qquad (15)$$

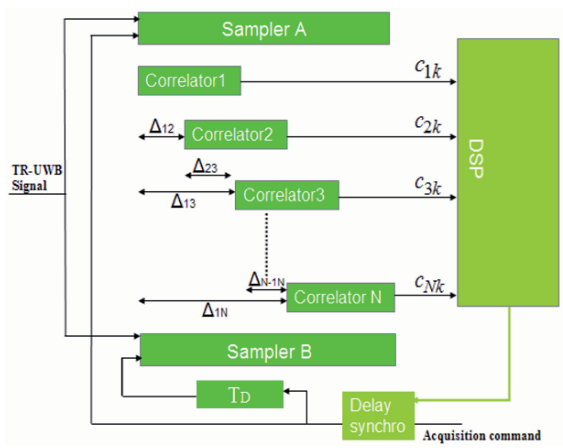


Figure 3. Receiver architecture with synchronization scheme for N correlators spaced in front of the samplers.

5 PERFORMANCE

With additive noise, local peaks can rise on correlation and the estimate of the maximum can be far from the good value. So it is interesting to evaluate the probability that the estimate point is in the good interval. To allow the compromise between the complexity and the performance it is also necessary to evaluate the loss in performance due to the sampling precision Δ .

5.1 *Correlation and noise model*

To evaluate the correlation and its noise, we have the shape of the received signal $d_r(t)$ in (2) made of a doublet of two delayed and spread pulses $g_r(t)$ (3) and an AWGN $n(t)$. The received doublet (or bit) energy is E_b . The noise is supposed limited to the signal wide band by an unit filter, so it has a narrow correlation function $R_n(\tau)$ less than Δ and a noise power equal to the variance of the sampled noise :

$$P_n = \sigma_n^2 = N_0/2 \qquad (16)$$

Assuming a correct waveform sampling period T_e , a doublet separation $T_D = N_D T_e$ larger than the channel spread (no IPI) and a correlation time $T_W = N_W T_e$ less than T_D , we can write the output of a correlator at the sampling time t as the sum of four terms :

$$c(t) = s^2(t) + g_r(t)n_2(t) + n_1(t)g_r(t) + n_1(t)n_2(t) \qquad (17)$$

The noiseless correlation is the signal part $s^2(t)$, it is the sum of all the paths present in the window duration T_W :

$$s^2(t) = \sum_{kT_e=t-T_W}^{t} g_r^2(k) \qquad (18)$$

With short pulses and following (4) the amplitude of the noiseless signal can be written as the collection of the energy of the paths multiplied by the energy of the emitted pulse. So we can write (18) as the product of the bit energy E_b at the receiver input and a coefficient $c_W(t)$ depending on the channel spread, the window length T_W and the sampling time t :

$$s^2(t) = X_c^2 \sum_{\tau_i \le T_W} \alpha_i^2 E_g = c_W(t) E_b \qquad (19)$$

If T_W exceed the channel spread (full rake) and if t samples the correlation maximum, then $s^2(t)$ gets its maximum value $E_b/2$.

The other terms in (17) are independent noises, the two first are Gaussian centered with variance σ_{gn}^2 :

$$g_r(t)n_i(t) = \sum_{kT_e=t-T_W}^{t} n(k)g_r(k) \qquad (20)$$

$$\sigma_{gn}^2 = X_c^2 \sum_{\tau_i \le T_W} \alpha_i^2 E_g N_0/2 = c_W(t) E_b P_n \qquad (21)$$

The last term is a sum of non gaussian RVs, from the central limit theorem it can be approximated by a Gaussian one with mean $m_{nn} = N_W R_n(T_D) \approx 0$ and variance σ_{nn}^2 :

$$n_1(t)n_2(t) = \sum_{kT_e=t-T_W}^{t} n(k)n(k-N_D) \tag{22}$$

$$\sigma_{nn}^2 = N_W\left(N_0/2\right)^2 = N_W P_n^2 \tag{23}$$

Finally the correlation point $c(t)$ can be approximated as a noiseless amplitude $c_W(t)E_b$, variable in time, affected by a gaussian centered noise $u(t)$ with variance $\sigma_u^2(t)$.

$$c(t) = c_W(t)E_b + u(t) \tag{24}$$

$$\sigma_u^2(t) = P_n\left(2c_W(t)E_b + N_W P_n\right) \tag{25}$$

5.2 *Probability of synchronization*

The synchronization starting time τ_1 can be any point within the symbol period $[0,T_s]$ with a uniform distribution. This interval is sampled by $M = KN$ points τ_m equally spaced by Δ. The probability of a correct synchronization is the probability that the estimated value $\hat{c}_{\max}$ is within an interval Δ around the exact maximum $c_{\max}$. Calling Δ_m the interval around the point τ_m and H_m the hypothesis that $\tau_{\max} \in \Delta_m$, the probability of a correct synchronization is then :

$$P_{sync} = \sum_{m=1}^{M} \Pr\left(c_m > c_j, \forall j \neq m / H_m\right)\Pr\left(H_m\right) \tag{26}$$

Due to the uniform distribution $\Pr\left(H_m\right) = 1/M$ and :

$$P_{sync} = \Pr\left(c_m > c_j, \forall j \neq m / H_m\right) \tag{27}$$

$$P_{sync} = \int_{c_m=-\infty}^{+\infty} \int_{c_j<c_m} \cdots \int p\left(c_1,\cdots,c_M / H_m\right)dc_1\cdots dc_M \tag{28}$$

The noise correlation is very short, we can assume independent variables and we can rewrite (28) :

$$P_{sync} = \int_{c_m=-\infty}^{+\infty} p\left(c_m / H_m\right) \int_{-\infty}^{c_m} p\left(c_1 / H_m\right) \int_{-\infty}^{c_m} \cdots \int_{-\infty}^{c_m} p\left(c_M / H_m\right)dc_1\cdots dc_M$$

We can note that the independence hypothesis is not always true since two successive correlators which give samples c_m and c_{m+1} of $c(t)$, can overlap (Fig.3) .

The laws $p\left(c_j / H_m\right)$ are known as gaussian using (24) and the sample amplitudes $c_j = c_W(\tau_j)E_b$ are fixed by the hypothesis H_m on a particular channel correlation. We can evaluate P_{sync} for individual channels and a mean value over these channels gives a good idea on the ability of synchronization for the CM1-4 models. A theoretical value can also be obtained using a mean correlation curve for each of these models; good approximations of $c_W(t)E_b$ are exponential values.

5.3 *Performance with synchronization*

Using the value $c(t)$ in (24) as detection sample gives, with a single bipolar chip, a fixed channel and a correlator window of size N_W, the error probability (see [15]) :

$$P_{e/c,W} = Q\left(\sqrt{\frac{E_b}{N_0}\frac{c_W(t)}{\left(1+\frac{N_W}{c_W(t)}\frac{N_0}{4E_b}\right)}}\right) \tag{29}$$

To improve the performance, we can use a smoothing to actively reduce the noise in the correlation outputs while the synchronization process. This mechanism is simply made with an increase of the preamble length by repeating the training sequence L times. The DSP calculates the smoothing correlation output c_{nks} :

$$c_{nks} = \sum_{i=1}^{L} c_{nk}(i) \tag{30}$$

The c_{nk} may have opposite polarity depending on the transmitted data. So to avoid destructive components the DSP must work with the absolute value of the outputs or the preamble sent could be made with same kind of bit 0 or 1.

6 SIMULATION

We use the optimal correlation window $T_W = 15\text{ns}$ evaluated in [15] and $\Delta_{1,N} = 12ns$ to simulate two TDSC architectures: with delta (Δ) = 6ns and 3ns, it gives architectures with N=3 and N=5 correlators respectively.

The analysis of the synchronization algorithm is performed for UWB channel models. The channel length reduces the data rate of the system ($T_s = 300\text{ns}$). The emitted doublet (bit) has $T_D = 150\text{ns}$ to avoid IPI and a number of 22 packets is transmitted. Each packet contains 500 bits in BPSK modulation. For smoothing we use L = 4, the same training sequence is send four times.

We simulate the TDSC receiver without frame or pulse averaging, and draws the BER and P_{sync} curves versus E_b/N_0 for each architecture. We use the IEEE indoor models [14] averaging over the 100 channels.

For each architecture with channel models CM1 & CM2, Fig.5 and Fig.7 show the simulated BER curves , and Fig. 4 and Fig. 6 show the probability of synchronization. We can see that with 3 correlators (delta (Δ) = 6ns) , the performance degradation is not remarkable but we have less synchronization accuracy. This effect comes from the small variation around the maximum of correlation and the channel spreading.

7 CONCLUSIONS

In this paper, a synchronization scheme for the TDSC principle has been presented. It uses a training sequence for fast correlation estimation. We have developed a theoretical expression to show the importance of the optimality of the

synchronization time. We also presented an analytical expression for the synchronization probability. A smoothing is used to enhance performance with an increase in the synchronization preamble length. The robustness of TDSC receiver to synchronization errors is shown, which reveals the existence of a tradeoff between performance and synchronization complexity. Simulation show that the use of 3 correlators is sufficient for the synchronization of the TDSC system. It ensures good performance with low complexity for silicon integration.

REFERENCES

[1] J.Kunisch and J.Pamp, "Measurement Results and Modeling aspects for the UWB Radio Channels", Proc. IEEE Conf. UWB Sys. and Tech., 2002, pp.19-23.

[2] M. Z. Win, R. A. Scholtz, "On the energy capture of ultra-wide bandwidth signals in dense multipath environments", IEEE Commun. Lett., vol. 2, Sep. 1998, pp. 245-247.

[3] A. Rabbachin, I. Oppermann, "Comparison of UWB transmitted reference schemes", Communications, IEE Proceedings. Vol 153, Issue 1, Feb. 2006, pp136-142

[4] R. Hoctor and H. Tomlinson, "Delay-hopped transmitted- reference RF communications," IEEE Conf. Ultra Wideband Systems and Techno., pp. 265-269, May 2002.

[5] Molisch, A. F. and Foerster, J.R. and Pendergrass, M., 2003, Channel Models for Ultrawideband Personal Area Networks. In Wireless Communications, Vol. 10, Issue: 6, pp: 14 – 21.

[6] N. Van Stralen, A. Dentinger, K. Welles, II., R. Gaus, R. Hoctor and H.Tomlinson, "Delay hopped transmitted-reference experimental results," IEEE Conf. Ultra Wideband Systems and Techno., pp. 93-98, May 2002.

[7] Dennis L. Goeckel and Qu Zhang, 2005, Slightly Frequency- Shifted Reference Ultra-wideband (UWB) Radio: TR-UWB without the Delay Element, IEEE MILCOM 2005.

[8] Yang Ni, 2005, "Récepteur UWB et procédé et système de transmission de données", French Patent GET/INT, No.0500886.

[9] M. Muller, Y. Ni, R. Lamberti, C. Saber, "A New Signal Detection Method for TR-UWB - By Time Delayed Sampling & Correlation (TDSC) " , WINSYS, Aug. 2006.

[10] Ning He and C. Tepedelenlioglu, "Performance analysis of noncoherent UWB receivers at different synchronization levels", Wireless Communications, IEEE Trans, Vol 5, N° 6, June 2006, pp 1266-1273

[11] L. Yang, G. Giannakis, A. Swami, "Noncoherent Ultra-Wideband (De)Modulation", Communications, IEEE Transactions on, Vol 55, N° 4, Apr 2007 pp 810-819

[12] C.Carbonelli and U. Mengali, "Synchronization algorithms for UWB signals" Communications, IEEE Transactions on, Volume 54, Issue 2, Feb. 2006, pp. 329-338

[13] S. Kleinfelder, "Gigahertz waveform sampling and digitization circuit design and implementation," IEEE Transactions on Nuclear Science, vol. 50, no. 4, Aug. 2003.

[14] J. Foerster, "Channel Modeling Sub-committee Report Final", IEEE P802.15-02/490r1-SG3a, Tech. Rep., Feb. 2003. http://ieee802.org/15/

[15] C. Saber, R. Lamberti, M. Muller, Y. Ni, F. Hirata-Flores. "Optimal Correlator Size for The TDSC-UWB Detection Method". Conf. Mobility 2007, 10-12 September 2007, Singapore.

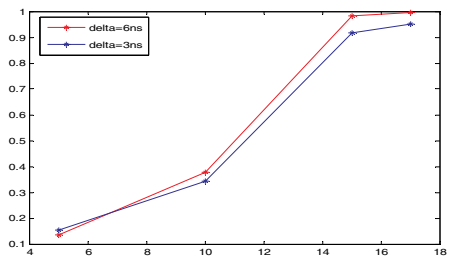


Fig.4 Comparison P_{sync} vs E_b / N_0 , delta=3ns, 6ns CM1

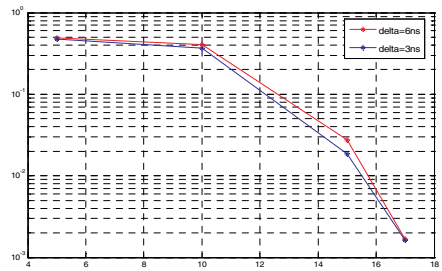


Fig. 5 Comparison BER vs E_b / N_0 , delta=3ns, 6ns CM1

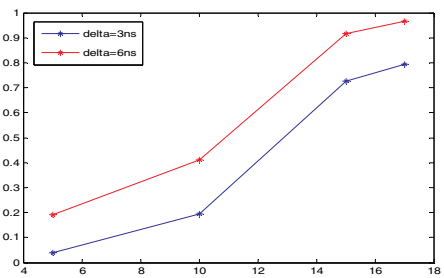


Fig.6 Comparison P_{sync} vs E_b / N_0 , delta=3ns, 6ns CM2

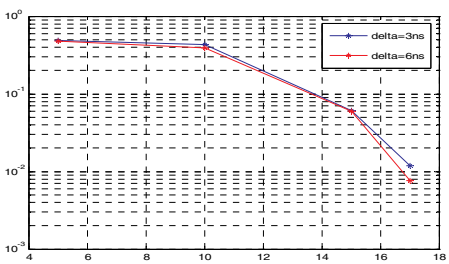


Fig. 7 Comparison BER vs E_b / N_0 , delta=3ns, 6ns CM2

An Efficient In-Network Event Detection Algorithm for Wireless Sensor Nodes

Chirakkal V. Easwaran
Department of Computer Science
State University of New York
New Paltz, NY 12561
easwaran@newpaltz.edu

ABSTRACT
In this paper we present an algorithm that allows a wireless sensor node to decide with reasonable error tolerance whether the node is within a region of interest or not. The algorithm requires a very limited amount of information exchange with neighboring nodes, making it very energy-efficient for resource-scarce sensor nodes. The sensitivity of event detection using this algorithm can be adjusted using a multiple-scale method. We show preliminary simulation results to demonstrate the effectiveness of the algorithms.

Categories and Subject Descriptors
I.6.5 [**Simulation and Modeling**]: Model Development – *Modeling Methodologies.*

General Terms
Algorithms, Measurement, Performance, Design.

Keywords
Sensor Networks, Geo-Spatial events.

1. INTRODUCTION

1.1 Sensor Networks
Wireless sensor networks (WSN) provide an architecture for inexpensive, ubiquitous data collection from the environment. Habitat monitoring, detection of chemicals in the environment and humidity and aridity monitoring in agricultural fields are examples of application areas that could take advantage of the evolving WSN hardware and software technology.

Research into the hardware and software aspects of sensor networks has evolved rapidly in recent years. Sensor platforms from companies such as Crossbow Inc., Telos and Dust Inc. have provided the impetus for hardware research in this area. A large number of research papers in recent years have also discussed various communication, data collection and data utilization aspects of sensor networks[1-2].

Paper presented at the *International Conference on Telecommunications and Networking, December 2007*.

Several recent papers ([5-8]) have discussed the use of sensor networks for detecting and monitoring of spatial information. Geo-spatial phenomena, which may generally be considered subsets of all spatial phenomena, are usually continuous in space and slowly-varying in time. In many cases, qualitative aspects of some region,, rather than the exact quantitative measurements, are more relevant in the monitoring of geo-spatial information. For example a distinction between a "hot" and "cold" region, according to some pre-determined criterion, is more relevant and sufficient than exact temperature readings of each spatial coordinate point in a geographic region. Similarly, identification of regions that have concentration of a chemical above a threshold T is more relevant than the actual measurement of chemical concentration at each spatial point. Such qualitative feature detection is generally termed "Event detection" in literature.

Massively deployed environmental and geo-spatial sensors are supposed to be inexpensive, energy and resource-constrained devices that may be used for multiple application scenarios. In the initially proposed architectures for sensor networks, these individual sensors gather data and transmit them to a master device with higher computational resources. The master device makes the necessary decisions based on the accumulated information from all the sensors in its reach. A problem with this scenario is the difficulty of scaling to bigger geographical regions and a large number of sensors, with the accompanying issues of communication costs and power requirements.

1.2 In-network Event Detection
An alternative architecture has been proposed in which the sensors themselves are enabled to make qualitative decisions based on data acquired from the environment. For example, a temperature-sensing node may be enabled to make a decision whether it is in a "hot" region or not, based on some pre-defined notion of being "hot". Such a process is termed *in-network* event detection. The capability for in-network event detection scales easily and cuts down un-necessary network communication, since generally each node needs to communicate with only nodes in its immediate neighborhood. Furthermore, only the sensors that detect an event need to communicate that event to a master device, thus reducing bandwidth consumption even further. In a typical application scenario, a grid of sensor nodes may be employed to measure soil moisture content in an agricultural

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 317–322.

field, and based on the localized aridity level, a sprinkler may be activated locally.

In this paper we present a new algorithm for in-network event detection by sensor nodes using a fixed-length coding mechanism that conserves transmission bandwidth of inter-node communication. The algorithm is noise-tolerant and provides high-quality in-network event detection capability for sensor networks.

The rest of this paper is organized as follows. In Section 2, we explore related work in the area. In Section 3, we propose a new algorithm for in-network event detection. In Section 4 we present preliminary results of our simulation using the proposed algorithm. Section 5 is a discussion of ongoing work and future directions of research.

2. RELATED WORK

Much of the current research on sensor networks focus on the retrieval of individual sensor node data. In [4], sensor networks are structured into triangular grids to support in-network query processing and qualitative information retrieval. There is also related work on sensor DBMSs that support complex queries and data aggregation [9].

Various aspects of in-network event and event boundary detection have been studied in a number of papers [3-8]. In [6], a Bayesian distributed algorithm is presented where event information is encoded into binary values and broadcast to neighboring nodes. It is proved in [6] that a majority-based decision scheme is optimal in these circumstances. However, the algorithm is not noise-tolerant, and requires location information from the nodes. Another type of algorithms makes use of moving means and medians of neighborhood sensor values to make estimations of event occurrence [3, 5].

A common characteristic of many of the proposed event detection algorithms is the direct relationship between communication and energy costs and detection accuracy. Essentially, you gain one at the expense of the other. Another aspect of sensor networks that constrains the development of algorithms is the simplicity of sensor hardware and software.

A relevant issue is whether the location of a sensor in space needs to be known to itself as well as to its neighbors. In most previous work [3-8] on event detection it is assumed that the sensors are location-aware either through deployment characteristics, using an RF beam, or through GPS systems. In reality, given that the sensors are expected to be very inexpensive and expendable devices, it is unlikely that these devices will be equipped with such functionality. Therefore, it is important to develop algorithms that do not depend on sensors being aware of their spatial location.

In [5], an algorithm has been proposed that is noise-tolerant, scalable and does not depend on the spatial localization of individual sensors. The algorithm, named NED (Noise Tolerant Event and Event Boundary Detection), assumes that all sensors are aware of a threshold, *T*, that defines the event in accordance with equation (1). All sensors are assumed to have the same Normal noise distribution $N(\mu,\sigma)$. Since roughly 95% of the mass of a normal density function $N(\mu,\sigma)$ falls within the range $(\mu-1.96\sigma,\mu+1.96\sigma)$, it is assumed that a sensor's detection of a reading $Y \geq T + 1.96\sigma$ signals a *significant event* while a reading of $Y \leq T - 1.96\sigma$ signals a *significant non-event*. According to the NED scheme, in both these situations, the sensor broadcasts two bits to inform its neighbors of its readings – the first bit is 0 signaling a significant event, and the second bit is 1 for a significant event or 0 for a significant non-event. In the complementary case where the sensor readings are in the interval $(T-1.96\sigma, T+1.96\sigma)$, the first bit is 1, and an additional 32 bits are used to convey the original sensor readings as a float value.

One difficulty with the above algorithm is the need for variable number of bits for transmission depending on sensor measurements, either 2 bits or 33 bits. In addition, the algorithm doesn't allow for dynamic scaling of sensitivity of nodes to suit particular needs.

In view of these observations, it is highly desirable to devise improved algorithms for efficient event detection, keeping in mind that such algorithms must be simple to implement and cost-efficient, easily scalable, low in network communication needs, as well as energy efficient for prolonged unattended use. This paper is a contribution towards these goals.

3. EVENT DETECTION ALGORITHM

3.1 Preliminaries

3.1.1 Definition of an Event.

As an approximation, we model the event space as 2-dimensional, which is acceptable for many geo-spatial applications. Inexpensive sensors are deployed in this space in an arbitrary fashion, and each sensor is capable of measuring a scalar quantity $Y_i(x_i,y_i)$ at the spatial position (x_i,y_i). The phenomenon of interest is assumed to be stationary. Each sensor can transmit to a short distance, and can receive transmissions from neighbors within a short radial neighborhood N_i. Each sensor measurement is supposed to include a random noise, which is generally modeled as Normally distributed, $N(0,\sigma)$. An *Event*, $E(x_i,y_i)$, at a sensor node at (x_i,y_i) may be modeled in multiple ways depending on the need of each application. In the model considered in this paper, we say that an event $E(x_i,y_i)$ occurred when the sensor reading at that location exceeds (or is less than, as the case may be) a threshold value *T*. For the sake of definiteness, we define

$$E(x_i,y_i) = \begin{cases} \text{True, if } Y_i \geq T \\ \text{False, otherwise} \end{cases} \quad (1)$$

where *T* is some pre-fixed threshold of the measured scalar variable. For example, $E(x_i,y_i)$ might signify the event that the temperature at (x_i,y_i) is greater than T= 20˚C at a given time. The question then is, how can an individual sensor determine, with reasonable confidence, the value of *E* at its location? In other words how can the sensor decide if it is *in event E* (*E* is true) or not (*E* is false) ?

Another type of "event" commonly defined in sensor network literature, for example [11], is based on detecting significant changes in sensor readings over spatial and temporal domains.

This definition of an event as a change of state is not used in this paper.

3.2 A new transmission coding algorithm

The encoding algorithm we propose is inspired by the one in [5], but incorporates substantial improvements. Our main contributions are:

(1) to reduce the number of bits transmitted by each node, thereby conserving network bandwidth utilization and reducing energy consumption by the nodes

(2) to make the transmission scheme uniform for all nodes with a constant length encoding scheme, and

(3) to increase the quality of event detection.

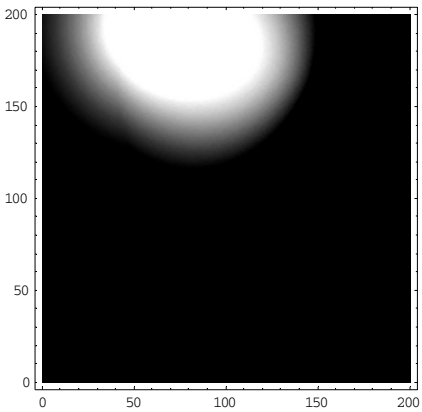


Figure 1. An event region is shown in white

3.2.1 Assumptions

We assume that

(i) Each sensor is calibrated with the threshold value T for the purpose of event detection according to Equation 1. This threshold is assumed to be the same constant for all sensors.

(ii) Each sensor measurement has a noise component that is normally distributed with mean 0 and variance σ^2.

3.3 The encoding and decoding scheme

3.3.1 Encoding algorithm

In order for a sensor to decide if it is detecting an event or not as defined by (1), it needs to consider its own sensor readings as well as those transmitted by neighbors. This in turn requires that sensors be able to digitize and transmit their sensor readings, and receive and decode the transmitted by other neighboring nodes. In some proposed schemes to encode sensor readings, such as in [6], the events are binary encoded into 1 or 0 for event and non-event respectively. In other schemes, such as those in [3], the readings are encoded into 32-bit floating point values. While the first scheme is imprecise, it is quite efficient because of the small amount of data that the networks need to handle. The second types of schemes increase processing and network burden considerably to achieve higher accuracy. The algorithm in [5] presents a mixture of binary and floating point coding scheme depending on the value of the sensor's scalar measurement.

In the new encoding algorithm we propose, each sensor reading is digitized into a six-bit value. As shown in detail below, two bits specify the interval around the threshold T that the sensor reading lies and the other four bits quantify the actual reading. It must be emphasized that the choice of *six* bits is somewhat arbitrary, and can be changed to any number of bits depending on the available computational resources and the need for accuracy. Higher number of bits would provide more accuracy at the cost of enhanced use of resources. The six-bit scheme we use in the following discussion serves as an example of the underlying concept.

Coding and Transmission Algorithm

Input: Y_i : the sensor reading.

T : Threshold value, fixed.

σ : The standard deviation of measurement error, assumed constant for all nodes.

1. Compute $T + 2\sigma$ and $T - 2\sigma$.

2.(a) If $Y_i \geq T + 2\sigma$, Compute $\left\lfloor \frac{Y_i - T}{\sigma} \right\rfloor \equiv \alpha$. (Note that $\alpha \geq 0$.)

If $\alpha \leqslant 15$,

encode α with 4 bits, say $b_1\ b_2\ b_3\ b_4$, and transmit **1 1 b_1 b_2 b_3 b_4** ;

Else

transmit **1 1 1 1 1 1**

End

2.(b) If $Y_i \leq T - 2\sigma$, Compute $\left\lfloor \frac{T - Y_i}{\sigma} \right\rfloor \equiv \alpha$. (Note that $\alpha \geq 0$.)

If $\alpha \leqslant 15$,

encode α with 4 bits, say $b_1 b_2 b_3 b_4$, and transmit **1 0 b_1 b_2 b_3 b_4**;

Else

transmit **1 0 1 1 1 1**

End

2.(c) If $T \leq Y_i \leq T + 2\sigma$, Compute $\left\lfloor \frac{Y_i - T}{2\sigma} \times 15 \right\rfloor \equiv \alpha$. (Note that $0 \leq \alpha \leq 15$.)

Encode α into a 4-bit binary number, $b_1 b_2 b_3 b_4$, and transmit **0 1 b_1 b_2 b_3 b_4**

End

2.(d) If $T - 2\sigma \le Y_i \le T$, Compute $\left\lfloor \frac{T - Y_i}{2\sigma} \times 15 \right\rfloor \equiv \alpha$. (Note that $0 \le \alpha \le 15$.)

Encode α into a 4-bit binary number,

$b_1\, b_2\, b_3\, b_4$, and transmit **0 0 b_1 b_2 b_3 b_4** .

End

The idea behind the encoding scheme is graphically shown in Figure 2.

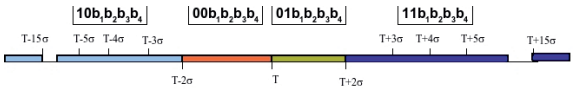


Figure 2. The encoding scheme

The first two bits (the most significant two bits) in the six-bit transmission of the sensor reading determine which of the four regions the value Y_i lies in:

First two bits	Meaning
00	$T - 2\sigma \le Y_i \le T$
01	$T \le Y_i \le T + 2\sigma$
10	$Y_i \le T - 2\sigma$
11	$Y_i \ge T + 2\sigma$

The remaining four bits are then used to encode the value of the sensor readings. See the example in Table 1 below.

Heuristically, if a sensor reading is within two standard deviations of the threshold, a finer scale is used to quantify and transmit the reading, using 4 bits to represent its scaled value (scaling as shown in steps 2(c) and 2(d) above). For readings beyond two standard deviations, a coarser resolution is used (4 bits to encode values in $(T + 2\sigma,\ T + 15\sigma)$ and $(T - 15\sigma,\ T - 2\sigma)$). Obviously, one can increase the precision of these approximate representations of the floating point values by increasing the number of bits chosen to transmit.

The interval ($T - 2\sigma$, $T + 2\sigma$) is a convenient representation of a slightly more than 95% confidence interval for the noisy sensor readings.

Observe that this encoding scheme provides a finer resolution of transmitted values in the interval ($T - 2\sigma$, $T + 2\sigma$), compared to the values outside this interval. This is very desirable because the sensor readings within two standard deviations of the mean have a higher chance of containing noise components. This enhancement in resolution in our algorithm considerably improves the accuracy of event detection.

Another significant advantage of the encoding presented here is that all sensor nodes use the same number of bits (six, in this case) for transmitting their readings. This is in contrast with the scheme in [4] that uses a two or a 33-bit transmission depending on the sensor reading and the pre-determined threshold T (See section 3.1.3).

3.3.2 Decoding

Decoding the 6-bit transmission is straightforward, working backward from the steps 1 through 2(d) in the previous section. The first two bits locate the region where the sensor values fall, and the next four bits provide a finer estimate of the values.

For illustration, suppose T=1, σ =0.5. Then the encoding of a few different sensor values and their decoding is shown below:

Table 1. Illustration of the encoding/decoding scheme (T=1, σ =0.5)

Sensor reading, Y_i	**Encoding**	**Decoded values**
7.3	111100	7
-4.3	101010	-4
1.75	011011	1.73
0.73	000100	0.733

3.4 Event detection

The event detection algorithm we use is based on the *moving means*. For each sensor node, the decoded sensor readings from its neighboring nodes, including its own readings, is averaged and if the computed mean is above the threshold, the sensor reports an event in accordance with equation (1).

4. SIMULATION RESULTS

Results from simulations using our encoding algorithm show very accurate prediction of event regions. We simulated geospatial events, similar to that in Figure 1, using an image with grey-scale pixel values (in the range 0-255, 0 being black). A pixel value greater than a threshold, (say a grey-scale pixel value of 60) constituted an event. We used a 200x200 pixel grid with sensors distributed at selected grid points. To estimate the *quality of event detection*, we used the metric

$$\text{Quality} = \frac{\text{Number of sensors correctly reporting the event}}{\text{Actual number of sensors in the event}} \times 100$$

We ran the current algorithm and the NED algorithm [5] a number of times over different scenarios including the one in Figure 1, to calculate the average event detection quality. The result from running the simulation with uniform regular distribution of sensors in the field and fixed threshold pixel value (60) is shown in Figure 3. Clearly our algorithm shows improved performance over that in [5] for wider noise distribution levels. We explain this as due to the higher accuracy achieved for each node's transmission.

Figure 4. shows the average detection quality for a fixed noise level (σ = .05) and threshold (T=60) as a function of sensor

density in the region. The advantage of our algorithm for *sparser* node distribution is obvious from this figure.

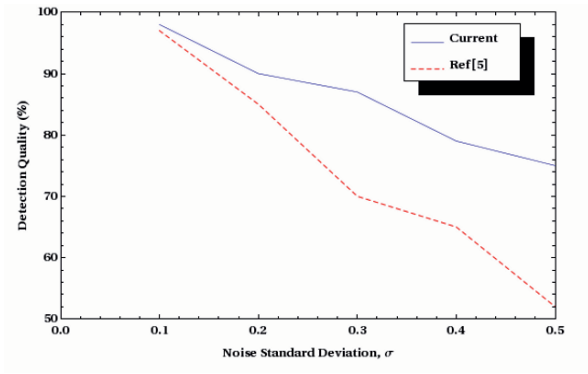


Figure 3. Event detection quality vs. noise SD

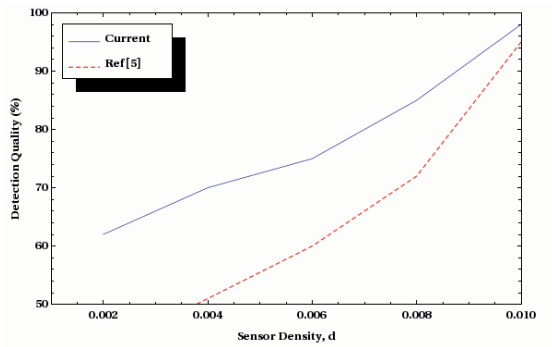


Figure 4. Event detection quality as a function of sensor density

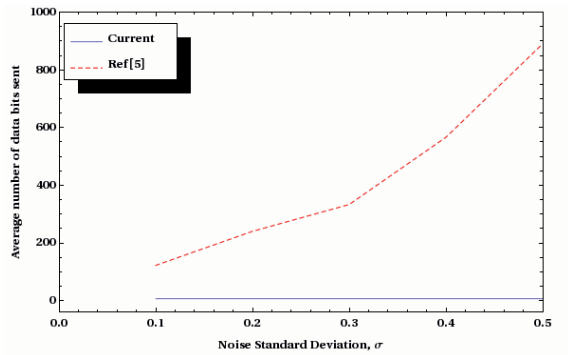


Figure 5. Average amount of data sent per node

In Figure 5 we show the considerable reduction in the number of transmitted bits using our algorithm over NED. This result is to be expected since our algorithm uses a fixed, 6-bit encoding for all transmissions, while NED uses 2 bits for "clearly" significant events beyond two standard deviations of the threshold and 33 bits (See section 3.1.3) for "noisy" measured values within two standard deviations.

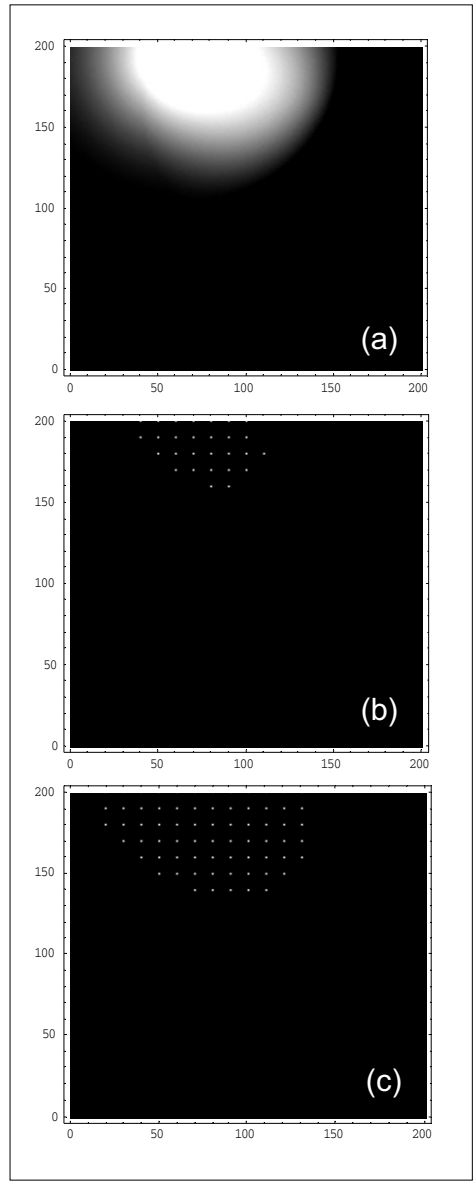


Figure 6. Visual Comparison of algorithms. An actual event is depicted in the white region in (a). Sensors reporting *In-Event* state using algorithm in [5] (white dots in (b)) and our algorithm (white dots in(c)) under identical test criteria.

Figure 6 shows qualitative comparison of sensors that picked up the event shown in Figure 1 using the algorithm in [5] and our

algorithm, respectively, for the same threshold and tolerance values. Note the significantly larger number of nodes reporting correct in-event state in Figure 6(c) compared to Figure 6(b). Our algorithm was capable of identifying a larger percentage of sensor nodes that were actually in the event, using far fewer encoding bits than in [5]. In test runs, this observation has been reproduced for different event region geometries.

5. DISCUSSION AND FUTURE WORK

We have presented a new data encoding and transmission algorithm for sensor nodes for in-network event detection. The scheme aggressively conserves bits, but still achieves a respectable level of accuracy. Unlike previous work, our algorithm does not require a differential encoding scheme – all sensor nodes transmit a fixed length (6 bits in our examples) message.

A significant achievement of our algorithm is the higher precision gained using only a six-bit transmission per event per node. In future work, we hope to take advantage of the spatially continuous and slowly-varying aspects of the underlying phenomena to achieve better event detection accuracy.

We are currently evaluating the accuracy and efficiency of the algorithm presented here against a number of existing algorithms [5, 6, 7, 8]. A thorough evaluation of the efficiency of various algorithms requires careful consideration of the metrics used for error and quality measurements in those schemes. In many cases, however, a direct comparison is infeasible because of the differences in their underlying assumptions. The results of our further evaluation will be reported in future conferences.

6. REFERENCES

[1] Culler, D. E. and Srivastava, M. "Overview of sensor networks", *Computer*, Vol. 37, no. 8, pp. 41–48, 2004.

[2] http://ceng.usc.edu/%7Eanrg/SensorNetBib.html (SensorNetBib- an online wireless sensor networks bibliography)

[3] Ding, M., Chen, D. , Xing, K and Cheng, X. "Localized fault-tolerant event boundary detection in sensor networks" in IEEE INFOCOM, vol. 2, pp. 902 – 913, IEEE, March 13-17, 2005.

[4] Duckham, M., Nittel, S. and Worboys, M. "Monitoring dynamic spatial fields using responsive geosensor networks" in Proceedings of the 13th ACM International Conference on Advanced in Geographic Information Systems (ACM-GIS05), pp. 51–60, ACM Press, 2005.

[5] Jin, G. and Nittel, S. "*NED:* An Efficient Noise-Tolerant Event and Event Boundary Detection Algorithm in Wireless Sensor Networks", *Proceedings of the 7th International Conference on Mobile Data Management, 2006*.

[6] Krishnamachari, B. and Iyengar, S. S. "Distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor networks", *IEEE Trans. Computers,* vol. 53, no. 3, pp. 241–250, 2004.

[7] Nowak, R. and Mitra, U. "Boundary estimation in sensor networks: Theory and methods", in 2nd International Workshop on Information Processing in Sensor Networks (IPSN) (Zhao, F. and Guibas, L.J., eds.), Vol. 2634, pp. 80–95, Springer Lecture Notes in Computer Science, 2003.

[8] Phani Kumar, A. V. U., Reddy, A. M. and Janakiram, D. "Distributed Collaboration for Event Detection in Wireless Sensor Networks", MPAC '05, November 28-December 2, 2005. Grenoble , France

[9] Madden, S., Franklin, M. J., Hellerstein, J. M. and Hong, W. "Tag: a tiny aggregation service for ad-hoc sensor networks", *SIGOPS Oper. Syst. Rev.*, vol. 36, no. SI, pp. 131–146, 2002.

[10] Gnawali, O., Greenstein,B., Jang, K-Y., Joki, A., Paek, J., Vieira, M., Estrin, D., Govindan, R., Kohler, E. "The Tenet Architecture for Tiered Sensor Networks", *SenSys '06*, November 2006.

[11] Xue, W., Luo, Q., Chen, L. and Liu,Y. "Contour Map Matching for Event Detection in Sensor Networks", Proceedings of *SIGMOD 2006*, Chicago, Illinois, USA.

Performance Evaluation of Distance Vector Routing Protocol on a Wireless Circular Model

D. C. Vasiliadis, G. E. Rizos, C. Vassilakis
Department of Computer Science and Technology
Faculty of Sciences and Technology
University of Peloponnese
GR-221 00 Tripolis
GREECE
dvas@uop.gr, georizos@uop.gr, costas@uop.gr

Abstract-In this paper, a wireless Circular Model over a generic distance-vector routing protocol is presented and analyzed. The performance of this model over the Distance Vector protocol, which is an implementation of Distributed Bellman-Ford algorithm, has been evaluated by using the simulation environment of NS-2. We conducted an extensive evaluation study for various mobility schemes in order to incorporate the behavior of nodes and the routing protocol in a real-life hotspot situation. In the test-bed model, while the number of source nodes was allowed to arbitrarily vary, there was exactly one destination node, closely modeling thus real-life situations where a single hotspot/access point exists. Finally, different constant bit rates (CBR) were used in order to estimate the throughput of receiving, dropping rates, the number of lost packets, as well as the average packet delay under various traffic conditions. This study is aimed to help wireless network designers in choosing the best suited routing protocols for their networks, through making explicit performance figures for common network setups.

1. INTRODUCTION

In the recent years, much research effort has been focusing on studying and improving the performance of routing protocols in Mobile Ad-hoc NETworks (MANETs) [1].

One significant classification scheme of routing protocols is based on the content of the routing tables. In the context of this classification scheme, two major classes of routing protocols can be identified, namely the Distance Vector (DV) and Link State (LS) protocols [2]. In a DV routing protocol such as the Routing Information Protocol (RIP) [3, 10], a vector is kept at each node, containing the cost (i.e. hop distance) and path (next hop) to all the destinations. In this protocol class, nodes exchange with each other a list (vector) of distances to the destinations, and each node maintains the routing table of the shortest paths to each known destination. DV protocols are generally known to suffer from slow route convergence and tendency of creating loops in mobile environments. On the other hand, a LS routing protocol such as the Open Shortest Path First (OSPF) protocol [4], overcomes the problem by maintaining the topology of the network (links state) at each router through periodical flooding of link information about its neighbours. Consequently, *medium or high mobility rates* entail frequent flooding. Unfortunately, this LS advertisement scheme generates larger routing *control overhead* than DV. In a network with population N, LS updating generates *routing overhead* in the order of $O(N^2)$. In large networks, the transmission of routing information will ultimately consume most of the *bandwidth*, at the expense of the useful bandwidth available to applications, which will be therefore blocked applications. Consequently, LS protocols are considered inappropriate for *bandwidth-limited* wireless ad-hoc networks, especially if these have medium or high mobility rates.

Another important classification scheme for routing protocols in Mobile Ad-hoc networks is based on the time that routing information is updated. Under this classification scheme, the classes of Proactive Routing Protocols (PRP) and Reactive Routing Protocols (RRP) [11] can be identified. Furthermore, a converged approach such as hybrid routing protocols considered.

Finally, another classification of routing protocols distinguishes them into source routing and hop-by-hop routing protocols. In source routing, the sources compute the complete path towards the destinations, leading to a loop-free routing protocol. On the other hand, in hop-by-hop routing, each intermediate node computes the next hop itself. Thus, the hop-by-hop routing protocols reduce the chance of *failed routes*, a parameter of crucial importance especially in mobile networks, which are more prone to the specific error type due to the fact that their topology changes much faster as compared to wired networks. Consequently, source routing protocols -such as the Dynamic Source Routing (DSR) [5]-

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 323–328.

allow intermediate nodes (and even overhearing nodes) to modify the route, adapting thus better to the nature of mobile networks. Most MANET routing protocols such as Optimized Link State Routing (OLSR) [6] and Ad-hoc On-demand Distance Vector (AODV) protocols [7] have adopted the strategy of hop-by-hop routing.

In this paper we study the performance of the Distance Vector routing protocol over various *mobility rates,* while the network topology follows the Circular Model. In particular, the following cases have been studied:

- *no mobility*
- *medium mobility*, where a medium amount of links fail and restore during the experiment period
- *high mobility*, where a large amount of the links fail and restore

We also, consider different traffic conditions and varying *packet size*, in order to study the effect of these parameters on the overall performance of the routing protocol.

The remainder of this paper is organized as follows: in section 2 we illustrate and analyze the test-bed wireless Circular Model over the Distance Vector protocol. Section 3 presents the results of our performance analysis, which has been conducted through simulation experiments, while section 4 provides the concluding remarks.

II. THE MODEL ANALYSIS

In this paper we analyze and evaluate the performance of DV routing protocol over a wireless Circular Model which is equivalent to the well-known ring topology used in standard Ethernet networks. It is worth noting that *traffic* in the presented model is routed along a route with the shortest number of hops from the *source node* to the *destination node*. This routing feature is closely resembles the strategy adopted by the AODV protocol. An important consideration of this model is that, although there are many *source nodes*, there is exactly one *destination node*, representing a single hotspot/access point.

This Circular Model simulates a number of *nodes* connected to a wireless hotspot, with the ability to have an alternative route established through the remaining *nodes*, when a *node-link* breaks. The network simulator NS-2 [8, 9] has been used for simulation study in order to investigate the potential problems in the terms of route maintenance and link failures. Unfortunately, NS-2 is very memory- and CPU-intensive, and therefore contributed to the small scale design and implementation of the models. Consequently, the slow simulation speed and large memory requirement of the NS-2 models prevented us from using larger networks at this evaluation study. The following Circular Model configuration was chosen to represent the behaviour of *nodes* and the routing protocol in a real-life hotspot situation:

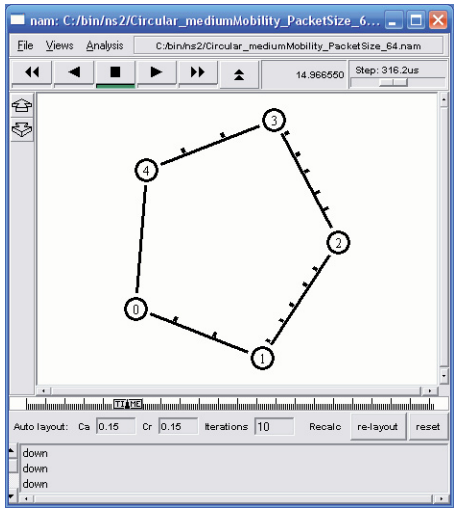


Fig. 1 Network Animation output for DV routing protocol on a *medium mobility* Circular Model under a stability of all *node-links*

- A five *node* Circular Model, consisting of four *source nodes* (0, 1, 3, and 4) all sending packets to the *destination node* 2, representing a single hotspot/access point. All links were set to a speed of 2 Mbps, with *propagation delay* 10 ms.
- The constant bit rate (CBR) *traffic* was used in the simulation by an *interval* of 0.004 sec (or 250 packets per second).
- The *packet size* was fixed either to 64 Bytes or to 512 Bytes for *low* and *high traffic* conditions respectively. In a slight variation the measurements presented in [13], the effect that different *packet size* may have on the performance will be investigated.
- The route with the smallest number of hops was being taken using the Distance Vector protocol which is an implementation of Distributed Bellman-Ford algorithm [12].
- The *simulation time* was adjusted to 60 sec.
- Three different *mobility* settings were implemented. In a *no mobility* setup all *node-links* were stable during the *simulation time*. In a *medium mobility* setup, a medium amount of links fail and restore during the *simulation time* (8 fails of a total duration 10 sec), while in the *high mobility* setup a large amount of links fail and restore (18 fails of a total duration 20 sec)

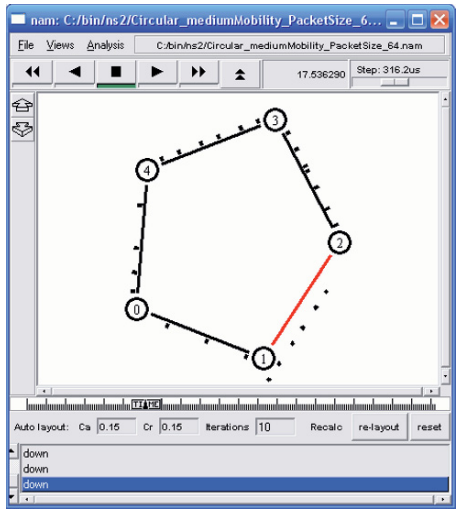


Fig. 2 Network Animation output for DV routing protocol on a *medium mobility* Circular Model under a *node-link* goes down

As illustrated in figure 1, the shortest route was taken by each *source node* (0, 1, 3, and 4) to the *destination node* 2. The routes were: 0→1→2, 1→2, 3→2, and 4→3→2. In figure 2, the link between *nodes* 1->2 was non-operational for a certain *time period* due to a *mobility* factor. During this period, the route for *node* 1 was changed from a single hop (1→2) to 1→0→4→3→2, while the route for *node* 0 was changed from 0→1→2, to 0→4→3→2. It is also seen at Network Animation output in figure 2 that a number of packets were dropped at the *node-link* 1→2, during a link failure.

III. PERFORMANCE AND SIMULATION RESULTS

The performance of the Distance Vector routing protocol over a wireless Circular Model under various *mobility* and *traffic* conditions has been evaluated by the NS-2 simulator [8, 9]. All simulations used a generic distance-vector routing protocol agent, as described by NS-2 manual [9]. The implementation sends periodic route updates every *advertInterval*. This variable is a configurable parameter that remains constant throughout the simulation, and its value was set at 2 seconds. In addition to periodic updates, each agent also sends triggered updates; it does this whenever the forwarding tables in the node change. This occurs either due to changes in the topology, or because an agent at the node received a route update, and recomputed and installed new routes. Each agent employs the split horizon with poisoned reverse mechanisms to advertise its routes to adjacent peers. "Split horizon" is the mechanism by which an agent will not advertise the route to a destination out of the interface that it is using to reach that destination. In a "Split horizon with poisoned reverse" mechanism, the agent will advertise that route out of that interface with a metric of infinity. Each DV agent uses a *preference_* of 120. The value is determined by the class variable of the same name. Each agent uses the class variable INFINITY (set at 32) to determine the validity of a route.

The following performance and operation metrics were collected:

1. *Number of dropped packets*. This metric was collected for all node pairs (*X, Y*), where *Y* is the sending node and *X* is the *receive/drop node.*
2. *Number of lost packets*. This metric was collected for all node pairs (X, Y), where *Y* is the sending node and *X* is the node that the packet was addressed to.
3. *Throughput of receiving bits*. This metric was collected for the *destination node* and the evolution of its value along the simulation time axis was recorded.
4. *Rate* of dropping bits at *receive and drop nodes* vs. *simulation time*.
5. *End to End Packet Delays*. The metric was collected and cumulative distribution diagrams were created to concisely present the effects of mobility and network load to this performance parameter.

Circular Model ***Low traffic*** **(64 Byte packets)**	**Total** ***Packets Generated***	**Total** ***Packets Dropped***	**Total** ***Packets Lost***	***Average Delay*** **(sec)**
No mobility	59330	0	0	0.015
Medium mobility	59455	34	32	0.016
High mobility	59613	73	72	0.018

TABLE 1: Packet information under *low traffic* conditions

Circular Model ***High traffic*** **(512 Byte packets)**	**Total** ***Packets Generated***	**Total** ***Packets Dropped***	**Total** ***Packets Lost***	***Average Delay*** **(sec)**
No mobility	59330	0	0	0.019
Medium mobility	59455	3497	2405	0.045
High mobility	59613	7851	4652	0.056

TABLE 2: Packet information under *high traffic* conditions

Tables 1 and 2 depict packet information regarding total packets which were generated, dropped and lost at *low* and *high traffic* conditions respectively, using the DV routing protocol over various *mobility* schemes for the circular model nodes. Metrics for the *average packet delay* are also shown, and it follows that the average packet delay is greater at *high mobility* setups. Furthermore, the *average packet delay* deteriorates considerably under *high traffic* conditions at both *medium* and *high mobility* configurations (181% and 211%, respectively), while in the "no mobility" setup the deterioration does exist, but is much less (26,6%).

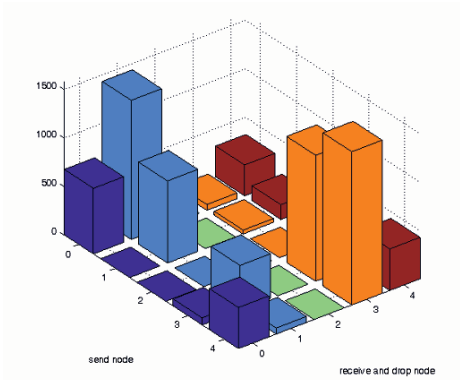


Fig. 3 *Numbers* of dropped packets at all *nodes X*: receive and drop *node Y*: send node

The graph in Figure 3 illustrates the number of dropped packets for all *node* pairs (*X, Y*) where *X* is the *node* that has received and subsequently dropped the packet, and *Y* is the original packet source. The metrics presented in these figures correspond to the circular model network setup described in section 2, which uses the DV routing protocol and exhibits high *node mobility* and high *packet rate* (0.5 Mbps per source). We may notice here that *node* 2 has no dropped packets at all, since packets having reached this *node* are considered to have reached their destination. *Nodes* neighbouring with *node* 2 (nodes 1 and 3) exhibit a higher number of dropped packets, since when their communication link to *node* 2 fails (*links* 1→2 and 3→2, respectively), this affects not only packets originating from these *nodes*, but for *nodes* 0 and 4, which use *nodes* 1 and 3 as intermediate hops for reaching *node* 2. Figure 3 illustrates the dropped packets at each *source node* (0, 1, 3, and 4), while figure 4 represents the lost packets at the *destination node* 2.

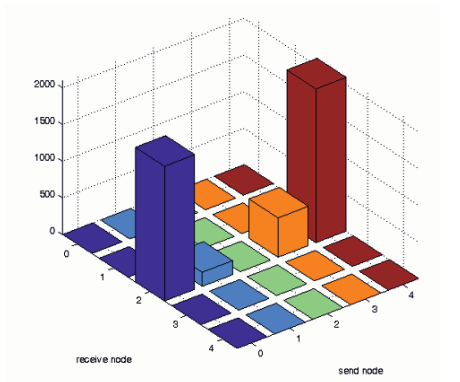


Fig. 4 *Numbers* of lost packets at all the *nodes X*: send *node Y*: receive node

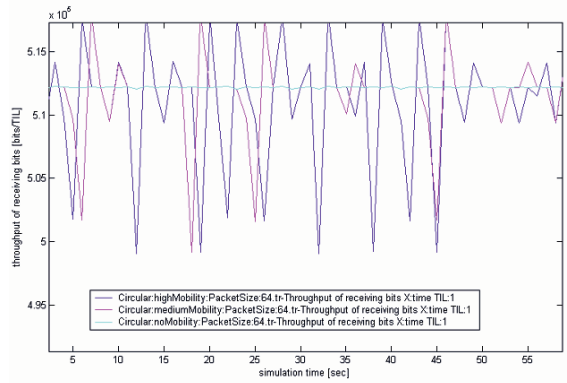


Fig.5 *Throughput* of receiving bits at *destination node* 2 vs. *simulation time* under *low traffic* conditions

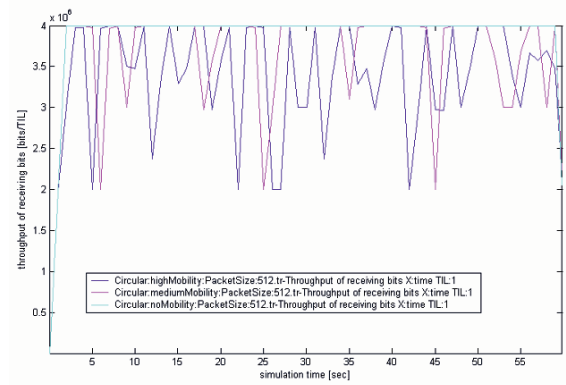


Fig.6 *Throughput* of receiving bits at *destination node* 2 vs. *simulation time* under *high traffic* conditions

Figures 5 and 6 depict the *average throughput* of receiving packets at the *destination node* 2 expressed in bps (bits per second) in the context of *low* and *high traffic* conditions respectively using the DV routing protocol. In each diagram, metrics for all mobility settings (high, medium and no mobility) are shown. The sharp performance drops that can be noticed for high and medium mobility setups are owing to link failures, as a result of mobility; this is also the reason that the "no mobility" setup does not exhibit such behaviour. It is worth noting that the *throughput* of receiving packets deteriorates slightly at *low* source transmission rates, under both *medium* and *high mobility* configurations. On the other hand, the *throughput* of receiving packets deteriorates considerably, under both *medium* and *high mobility* setups, when the source transmission rate becomes high. As can be seen from the diagrams, this difference can be attributed to the fact that in the low source *transmission rate*, the network has enough capacity to serve both the "regular" communication and the retransmitted (or rerouted) packets after a link failure. This explains the *throughput* spikes

following the sharp drops in figure 5, and these spikes partially compensate for the *bandwidth* lost due to link failures. On the contrary, when the source *transmission rate* is high, the network appears to not have adequate *bandwidth* to serve both regular communication and packet retransmissions/reroutings: in figure 6, the maximum *receiving throughput* observed is that of the "*no mobility*" setup, and no "spikes" above that limit are observed.

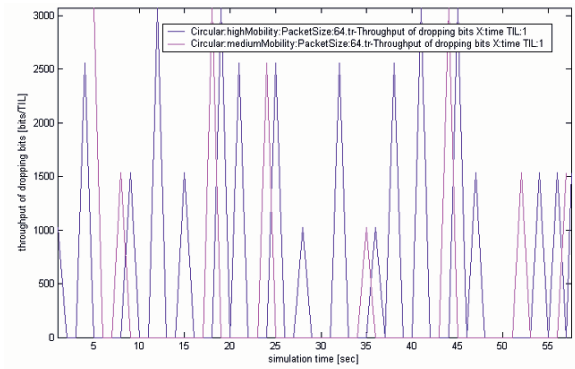


Fig.7 *Rate* of dropping bits vs. *simulation time* under *low traffic* conditions

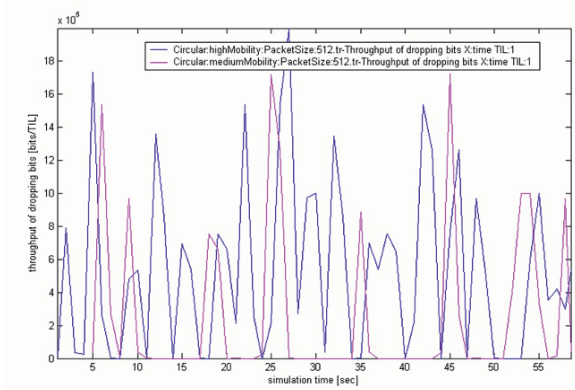


Fig.8 *Rate* of dropping bits vs. *simulation time* under *high traffic* conditions

Figures 7 and 8 represent the *cumulative rate* of dropping bits at all *receive* and *drop nodes,* under *low* and *high traffic* conditions respectively. It is noticed that the rate of dropping packets ranged from negligible to tolerable at *low traffic* conditions, under both *medium* and *high mobility* setups. On the other hand, the *cumulative rate* of dropping packets increases considerably in the case of *high traffic* configurations.

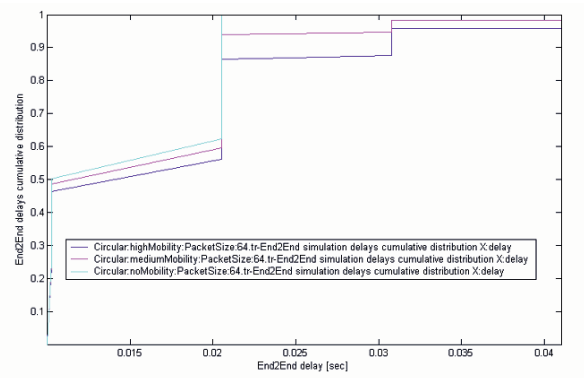


Fig. 9 *End to End Simulation Delays* vs. *Cumulative Distribution* under *low traffic* conditions

Finally figures 9 and 10 illustrate the *End to End Simulation Delays* vs. *Cumulative Distribution* using the DV routing protocol on various mobility patterns. Under the low traffic scheme and the "no mobility" setup we notice that packet service time effectively falls into two areas, the first one ranging from 0.1 to 0.2 sec and effectively corresponding to packets needing one hop to reach their destination (approximately 50% of the overall number of packets), while the second area corresponds to packets needing two hops to reach their destination. The variation in service times in the first area can be attributed to queuing delays at the source node, for the cases that the link is occupied by forwarding packets originating from other nodes to their destination when a new packet is generated. For *medium* and *high mobility* schemes, a third area is introduced to accommodate service time for packets needing to be retransmitted and/or following lengthier routes, due to communication link failures.

In the case of high traffic condition, the end to end delay presents a small increase under the "*no mobility*" scheme, mostly owing to the increase of queuing delays, since the probability that some link is occupied is greater than in the case of low traffic. For *high* and *medium mobility* patterns the end to end delay increases considerably, since –due to link failures owing to *mobility*– (a) packets follow lengthier routes and (b) the network appears not to have adequate capacity to effectively serve packets that need to be retransmitted and/or follow alternative routes.

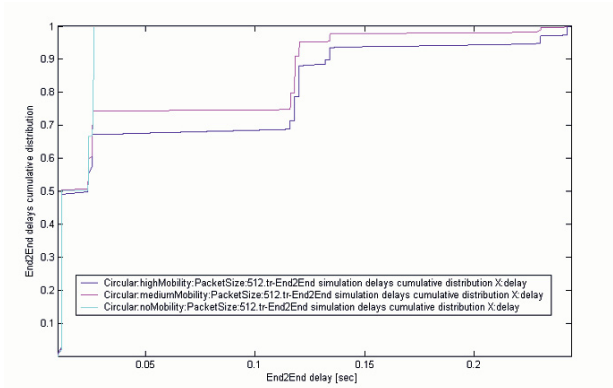


Fig. 10 *End to End Simulation Delays* vs. *Cumulative Distribution* under *high traffic* conditions

IV. CONCLUSIONS AND FUTURE WORK

In this paper an extensive performance evaluation of a wireless Circular Model over the Distance Vector routing protocol conducted under various *mobility* schemes. We observed that the *throughput* of receiving packets deteriorated slightly under low *traffic* conditions, at both medium and high *mobility* configurations. On the other hand, under high *traffic* conditions, the *throughput* of receiving packets deteriorated considerably, due to the flooding of sending packets at the last hops of the alternative *routing path* during a certain *time period* when a *link* between two *nodes* was inoperative due to the mobility factor. Moreover, the *average packet delay* was incremented dramatically at both *medium* and *high mobility* setups under *high traffic* conditions. The performance analysis is also uniformly applied to several representative networks such IEEE 802.11b and IEEE 802.11g providing a basis for fair comparison of *traffic* parameters and *mobility* conditions under tolerable *bandwidths* of drop and lost packets with acceptable *packet delays*.

Future work will include studying configurations with varying node populations, setups with multiple hotspots as well as setups where *mobility* does not lead only to communication link failures, but to alteration of route paths (e.g. when a node moves closer to the hotspot, it may directly transmit packets to it instead of using an intermediate hop). Correlation of *network loads* and *packet rates* to application classes (for instance web browsing, file transfer, streaming media and so forth) and study of network performance different under application usage patterns will be also considered.

V. REFERENCES

[1] S. Corson and J. Macker, "Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations", *RFC2501*, Jan. 1999.

[2] S. Keshav, "An Engineering Approach to Computer Networking: ATM networks, the Internet, and the telephone network", *Chapter 11, Addison Wesley Inc*., 1997.

[3] D. Pei, D. Massey, and L. Zhang. "Detection of Invalid announcements in RIP protocols", *IEEE Globecom 2003*, San Francisco, California, USA. December 2003.

[4] F.Y. Wang and F.S. Wu. "On the Vulnerability and Protection of OSPF Routing Protocol". *In Proceedings of the IEEE Seventh International Conference on Computer Communications and Networks*, Lafayette, LA, USA. Oct 12-15, 1998

[5] D.B. Johnson and D.A. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks," *Mobile Computing, edited by T. Imielinski and H. Korth, Chapter 5*, Kluwer Publishing Company, 1996, pp. 153-181.

[6] P. Jacquet, P. Muhlethaler, A. Qayyum, A. Laouiti, L. Viennot and T. Clausen, "Optimized Link State Routing Protocol," *draft-ietf-manet-olsr-05.txt, Internet Draft, IETF MANET Working Group*, Nov. 2000.

[7] Lu Yi, Wang Weichao, Zhong Yuhui, Bhargava Bharat, "Study of Distance Vector Routing Protocols for Mobile Ad Hoc Networks", *Proceedings of First IEEE International Conference on Pervasive Computing and Communications*, *(PerCom 2003)*, pp. 187-194, 23-26 March 2003.

[8] Fall K. and Varadhan K., ns notes and documentation, 1999, available from http://www-mash.cs.berkeley.edu/ns/.

[9] http://www.isi.edu/nsnam/ns/doc

[10] G. Malkin. RIP Version 2. *RFC 2453* (Standard). November 1998.

[11] Pham Peter P. and Perreau Sylvie, "Performance Analysis of Reactive Shortest Path and Multipath Routing Mechanism with Load Balance", *Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2003.* Volume 1, pp. 251 – 259, March 30 -April 3, 2003.

[12] Lynch, N.A.: Distributed Algorithms. Morgan Kaufmann, USA (1996)

[13] Perkins C, Royer E, Das S, Marina M. "Performance comparison of two on-demand routing protocols for ad hoc networks". IEEE Personal Communications, vol.8, no.1, Feb. 2001, pp.16-28.

Performance Evaluation of Mobile Ad-Hoc Routing Protocols

Eman Abdelfattah & Guinshin Liu

Computer Science Department
University of Bridgeport
Bridgeport, CT 06604

Abstract

In this paper we are considering the performance evaluation of three routing protocols in Mobile Ad-hoc Networks (MANETs). We have tested the effect of speed, TCP types, and maximum queue length for Destination Sequence Distance Vector (DSDV) protocol, Dynamic Source Routing (DSR), and Ad-hoc On Demand Distance Vector (AODV).

1. Mobile Ad-hoc Routing Protocols

Mobile Ad-hoc networks (MANETs) are the only choice when there is no infrastructure wireless network or it is too expensive to build one as the case in remote areas. They have the advantage of being quickly deployable.

One of the first ad-hoc wireless networks was the packet radio network started by ARPA in 1973[6]. It contained 138 nodes and used IP packets for data transport. This made an easy connection possible to the APRAnet, the starting point of today's Internet. A variant of distance vector routing protocol was used in this ad-hoc network.

Routing is needed to find a path between the source and the destination and to forward the packets even if the destination node is out of range of the source node. In ad-hoc networks, high mobility, low bandwidth, and limited computing capability characteristics of mobile hosts make routing of data is one of the most difficult issues.

In ad-hoc networks, we have asymmetric links which means that routing information collected for one direction is of almost no use for the other direction. Also, redundancy can not be controlled in ad-hoc networks. However, a high redundancy of links causes a large computational overhead for routing table updates. Furthermore, nodes may interfere and destroy each other. However, interference may help routing as a node can learn the topology with the help of packets it has overheard. Moreover, the greatest problem for routing protocols arises from the highly dynamic topology because the nodes must be able to keep up with the drastically and unpredictably changing in the network topology.

Destination Sequence Distance Vector (DSDV), Dynamic Source Routing (DSR), and Ad-hoc On Demand Distance Vector (AODV) are classified under flat ad-hoc routing protocols where all nodes play an equal role in routing and the addressing scheme is flat.

DSDV is a flat proactive (or table-driven) protocol. It sets up tables required for routing regardless of any traffic that would require routing functionality. An advantage of a proactive protocol is that it can give QoS guarantees related connection set-up, latency or other real-time requirements. A disadvantage of a proactive protocol is its overhead in lightly loaded networks. Periodic updates for the routing tables consume the batteries of mobile devices and generate a lot of unnecessary traffic (bandwidth).

Both DSR and AODV are flat reactive (or on-demand) protocols. They set up a path between the sender and the receiver only if a communication is waiting. An advantage of a reactive protocol is its scalability as long as there is only light traffic and low mobility. The disadvantages of these protocols are: (a) the initial search latency may degrade the performance of the interactive applications, (b) the quality of the path is unknown in advance, and (c) route caching mechanism is useless in high mobility networks as routes change frequently. [6]

1.1 Destination Sequence Distance Vector (DSDV)

DSDV routing protocol is an enhancement to distance vector routing for ad-hoc networks. A sequence number is used to tag each route. A route with higher sequence number is more favorable than a route with lower sequence number. However, if two routes have the same sequence number, the route with fewer hops is more favorable. In case of route failure, its hop number is set to infinity and its sequence number is increased to an odd number where even numbers are reserved only to connected paths.

1.2 Dynamic Source Routing (DSR)

DSR is an on-demand protocol. It is composed of route discovery and route maintenance. In route discovery, a node tries to discover a route to destination if it has to send data to this destination and there is currently no known route(s). A node broadcasts a route request (RREQ) with a unique identifier and the destination address as parameters. Any node that receives RREQ does the following [6]:

1. If it has already received the request, it drops the request packet.
2. If it recognizes its own address as the destination, then the request has reached its target.
3. Otherwise, the node appends its own address to a list of traversed hops in the packet and broadcasts this updated route request.

In route maintenance, a node is continuously sending packets via a route. The node has to make sure that the route is

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 329–334.

held upright. If a node detects problems with the current route, it has to find an alternative route.

1.3 Ad-hoc On Demand Distance Vector (AODV)

AODV uses different messages to discover and maintain links. In route discovery, a node broadcasts a Route Request (RREQ) to all nodes in the network till either the destination is reached or another node is found with a valid route entry for the destination whose associated sequence number is at least as great as that contained in RREQ. Then a Route Reply (RREP) is sent back to the source and the discovered route is made available. [5]

In route maintenance, when a node detects that a route to a neighbor node is not valid, it removes the routing entry and sends a Route Error (RERR) message to the active neighbors that use the route. This procedure is repeated at nodes that receive RERR messages. However, a source node that receives a RERR can reinitiate a RREQ message. Every route expires after a predetermined period of time. Sending a packet via a route will reset the associated expiration time.

1.4 Comparison of AODV and DSR

Table 1 presents a comparison of AODV and DSR Ad-hoc routing protocols [4].

	AODV	DSR
Type of routing protocol	Flat reactive	Flat reactive
Packet header overhead	Smaller than DSR	Larger than AODV which restricts the scalability of DSR
Route learning capability	Can learn less than DSR	Can learn more the AODV because DSR packets contain complete route information
Multiple path routing	Not applicable	Is applicable and can significantly improve the performance
Delivery of data packets	Higher percentage of data packets delivery on large networks (600 nodes) compared to DSR	Smaller percentage of data packets delivery on large networks (600 nodes) than AODV
Route information	Maintained in a distributed manner which is more suitable for MANETs	Maintained by the source
Security	Harder to secure than DSR	Easier to secure than AODV because the source can select a secure route based on the trust of nodes

Table 1 Comparison of AODV and DSR

2. Testing Environment

In this paper, we use NS simulator for simulating different routing protocols [1, 2]. NS simulator uses a visual tool called NAM. NAM is a Tcl/TK based animation tool for viewing network simulation traces and real world packet trace data. The design theory behind NAM was to create an animator that is able to read large animation data sets and be extensible enough so that it could be used in different network visualization situations.

The first step to use NAM is to produce the trace file. The trace file contains topology information, e.g., nodes, links, as well as packet traces. Usually, the trace file is generated by NS. During an NS simulation, a user can produce topology configurations, layout information, and packet traces using tracing events in NS. When the trace file is generated, it is ready to be animated by NAM. Upon startup, NAM will read the trace file, create topology, pop up a window, do layout if necessary. We have examined the following parameters for DSDV routing protocol:

1. Node 3 speed varies between 0 and 500
2. TCP protocol version: NewReno, Reno, Sack1, Vegas
3. Maximum queue length varies between 50 and 5

Our simulation scenario consists of tcl script that runs two TCP connections over a 4-nodes network over an area of a size of 500m over 400m as shown in Figure 1.

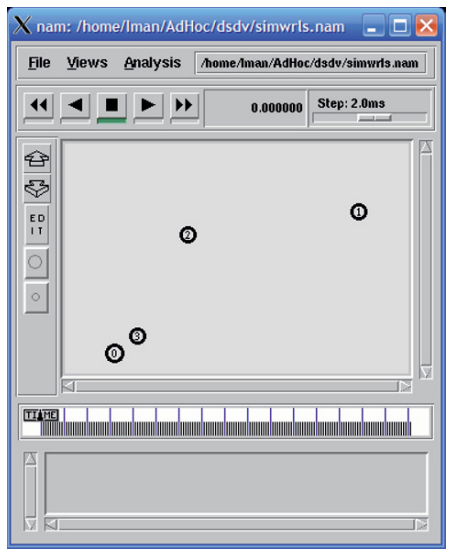


Figure 1: The topology of the network at time 0

3. DSDV Routing Protocol Simulation and Test Cases

3.1 Effect of Node's speed

For speed that ranges between 5 m/s to 100 m/s the performance of both sources of data does not change. The first

source of data has a congestion window size that increases rapidly in the beginning then behaves in a linear fashion. The second source of data has a constant congestion window size of 1 over the simulation period. For speeds higher than 100 m/s both data sources have a reasonable congestion window size over short period of the simulation time while it is 1 most of the time.

3.2 Effect of TCP protocol types

We have tested the NewReno, Reno, Sack1, and Vegas of TCP protocol. Each of NewReno and Reno exhibits similar performance. In these two test cases, the two sources of TCP connections do not show similar performance. For the first data source, the congestion window size increases rapidly in the beginning then behaves in a linear fashion. The second data source has a congestion window size of 1. For Sack1 protocol, the two sources of TCP connections do not show similar performance. The Vegas protocol shows a different performance. The congestion window size is very small fluctuates among values 1, 2, 3, 4, 5. Both sources exhibit the same behavior.

3.3 Effect of maximum queue length

We have tested the effect of the maximum queue length for a length that varies from 50 packets to 5 packets. In the first early test cases we start with a large window size of 50. In this case, the two sources of data do not show similar performance. For the first source, the congestion window size increases rapidly in the beginning then behaves in a linear fashion. The second source has a congestion window size of 1. As the queue size decreases, the window size never became stable. As we continue to decrease the size to 5, the window size keeps fluctuating due to the fact that packets are continuously dropped. This phenomenon can be attributed to the fact that the queue size is not large enough to keep a steady state of the congestion window size.

4. DSR Routing Protocol Simulation and Test Cases

4.1 Effect of Node's speed

In all the speeds used in the simulation tests the two sources of TCP connections show similar performance. The congestion window size increases rapidly in the beginning then behaves in a linear fashion.

4.2 Effect of TCP protocol types

We have tested the NewReno, Reno, Sack1, and Vegas of TCP protocol. Each of NewReno, Reno and Sack1 exhibits similar performance. In these three test cases, the two sources of TCP connections show similar performance. The congestion window size increases rapidly in the beginning then behaves in a linear fashion. The Vegas protocol shows a different performance. The congestion window size is very small fluctuates among values 1, 2, 3, 4, 5, and 6. Both sources exhibit the same fluctuating behavior.

4.3 Effect of maximum queue length

We have tested the effect of the maximum queue length for a length that varies from 50 packets to 5 packets. In the first early test cases we start with a large window size of 50. In this case, we have an excellent performance where the two sources of TCP connections show similar performance. The congestion window size increases rapidly in the beginning then behaves in a linear fashion. As the queue size decreases below 25, the window size never became stable. As we continue to decrease the size to 5, the window size keeps fluctuating due to the fact that packets are continuously dropped. This phenomenon can be attributed to the fact that the queue size is not large enough to keep a steady state of the congestion window size.

5. AODV Routing Protocol Simulation and Test Cases

5.1 Effect of Node's speed

In the first early test cases we start with node 3 at a speed of zero m/s. In this case the node is stationary and exhibits an excellent performance where the two sources of TCP connections show similar performance. The congestion window size increases rapidly in the beginning then behaves in a linear fashion. As the node speed increases, the congestion window size does not show such optimality. At a speed of 15 the two sources of TCP connections do not show similar performance. In one source, the congestion window size increases rapidly in the beginning then behaves in a linear fashion. There is a drop in the window size due to lost pockets. After the drop in the window size, there is a fast increase in the window size. The second source shows a small congestion window size due to the fact of the high speed it is moving with.

5.2 Effect of TCP protocol types

We have tested the NewReno, Reno, Sack1, and Vegas of TCP protocol. Each of NewReno, Reno and Sack1 exhibits similar performance. In these three test cases, the two sources of TCP connections show similar performance. The congestion window size increases rapidly in the beginning then behaves in a linear fashion. The Vegas protocol shows a different performance. The two sources of TCP connections have a small congestion window size.

5.3 Effect of maximum queue length

We have tested the effect of the maximum queue length for a length that varies from 50 packets to 5 packets. In the first early test cases we start with a large window size of 50. In this case, we have an excellent performance where the two sources of TCP connections show similar performance. The congestion window size increases rapidly in the beginning then behaves in a linear fashion. As the queue size decreases, the window size never became stable. As we continue to decrease the size to 5, the window size keeps fluctuating due to the fact that packets are continuously dropped. This phenomenon can be attributed

to the fact that the queue size is not large enough to keep a steady state of the congestion window size.

6. Results and Discussion

In this section we discuss the effect of speed of the moving node, maximum number of packets in the queue, and the TCP types on the Throughput of the mobile ad-hoc network.

To determine the total number of data packets that has been sent during simulation, the following commands were used

grep "^s" simple.tr | grep "_3_ AGT" > tcp2.tr
grep "^s" simple.tr | grep "_0_ AGT" > tcp3.tr

To determine the total number of data packets that has been received during simulation, the following command was used

grep "^r" simple.tr | grep "tcp" | grep "_1_ AGT" > tcp.tr

An example of a line in the output trace is

r 55.993815788 _1_ AGT --- 34 tcp 80 [**13a** 1 2 800] ------ [0:0 1:0 31 1] [0 0] 2 0

The bold data in that line is a hexadecimal number 13a which equals 314 in decimal, specifies the expected time in seconds to send this data over the wireless channel. To calculate the average delay in seconds to send data over the wireless channel, we used the following command

grep "\[**13a**" tcp.tr > tcp1.tr

However, # of records in tcp.tr equals # of records in tcp1.tr which means that all records in tcp.tr have a delay of 13a (314 in decimal) in seconds.

6.1 Throughput vs. Speed

Table 2 shows the data sent by each of the sources as well as the data received for each of the DSDV, DSR, AODV protocols versus speed.

	DSDV			DSR			AODV		
Speed	Sent S1	Sent S2	Received	Sent S1	Sent S2	Received	Sent S1	Sent S2	Received
0	3527	3524	7021	415	435	829	2568	2036	4587
5	5507	6	5483	571	558	1125	2527	1822	4340
10	5498	6	5481	1523	1524	3044	4130	1139	5236
15	5502	6	5479	1160	1177	2335	5309	544	5814
50	5190	161	5292	651	788	1436	6913	266	7129
100	5378	96	5427	285	303	563	4117	411	4474
200	1799	786	2539	247	263	484	3931	507	4401
300	1811	779	2541	230	251	467	4165	637	4758
500	1811	779	2541	215	234	443	4185	593	4731

Table 2: Data Sent and received under different protocols versus speed.

Table 3 shows the throughput for each protocol versus speed using the following calculation:

Throughput = data Received / (Data sent from S1 + Data sent from S2)

	Throughput		
Speed	DSDV	DSR	AODV
0	0.996	0.975	0.996
5	0.995	0.996	0.998
10	0.996	0.999	0.994
15	0.995	0.999	0.993
50	0.989	0.998	0.993
100	0.991	0.957	0.988
200	0.982	0.949	0.992
300	0.981	0.971	0.991
500	0.981	0.987	0.990

Table 3: Throughput versus speed for different Protocols.

Figure 2 shows the throughput for each protocol versus speed.

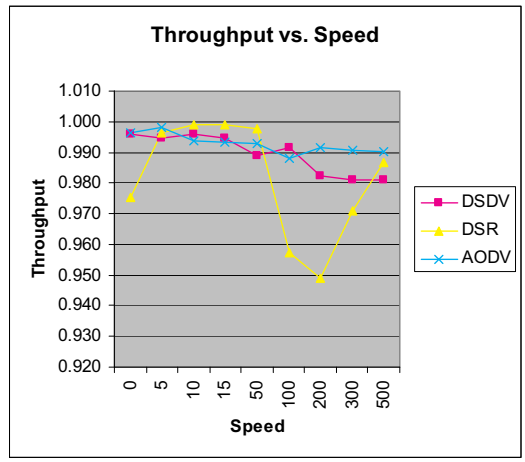


Figure 2: throughput for each protocol versus speed.

6.2 Throughput vs. TCP protocol types

Table 4 shows the data sent by each of the sources as well as the data received for each of the DSDV, DSR, AODV protocols versus TCP protocol types

TCP Type	DSDV			DSR			AODV		
	Sent S1	Sent S2	Received	Sent S1	Sent S2	Received	Sent S1	Sent S2	Received
Newreno	5507	6	5483	571	558	1125	2527	1822	4340
Reno	5507	6	5483	504	586	1085	2527	1822	4340
Sack1	3979	1197	5136	585	548	1130	2527	1822	4340
Vegas	3229	1726	4912	2431	1866	4295	2562	1931	4471

Table 4: Data Sent and received under different protocols versus TCP types.

Table 5 shows the throughput for each protocol versus TCP protocol types using the following calculation:

Throughput = data Received / (Data sent from S1 + Data sent from S2)

TCP Type	Throughput		
	DSDV	DSR	AODV
Newreno	0.995	0.996	0.998
Reno	0.995	0.995	0.998
Sack1	0.992	0.997	0.998
Vegas	0.991	1.000	0.995

Table 5: Throughput versus TCP protocol types for different Protocols.

Figure 3 shows the throughput for each of the DSDV, DSR, AODV protocols versus the TCP protocol types.

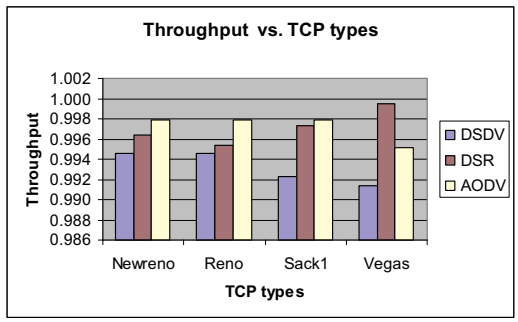


Figure 3: Throughput for each of the DSDV, DSR, AODV protocols versus TCP protocol types.

6.3 Throughput vs. Maximum Packets in queue

Table 6 shows the data sent by each of the sources as well as the data received for each of the DSDV, DSR, AODV protocols versus different values of maximum packets in the queue.

Max # of packets	DSDV			DSR			AODV		
	Sent S1	Sent S2	Received	Sent S1	Sent S2	Received	Sent S1	Sent S2	Received
50	5507	6	5483	571	558	1125	2527	1822	4340
40	5507	6	5483	571	558	1125	2527	1822	4340
25	5507	6	5483	2408	1925	4316	2851	1569	4368
20	5504	6	5482	585	575	1141	2394	2016	4317
15	4061	1527	5512	668	508	1151	2578	1838	4289
5	4023	1690	5428	2403	2113	4295	2882	1280	3811

Table 6: Data Sent and received for each of the DSDV, DSR, AODV protocols versus different values of maximum packets in the queue.

Table 7 shows the throughput for different protocols versus values of maximum packets in the queue using the following calculation:

Throughput = data Received / (Data sent from S1 + Data sent from S2)

Figure 4 shows the throughput for each of the DSDV, DSR, AODV protocols versus different values of maximum packets in the queue

Maximum packets	Throughput		
	DSDV	DSR	AODV
50	0.995	0.996	0.998
40	0.995	0.996	0.998
25	0.995	0.996	0.988
20	0.995	0.984	0.979
15	0.986	0.979	0.971
5	0.950	0.951	0.916

Table 7: Throughput versus TCP protocol types for different values of maximum packets in the queue.

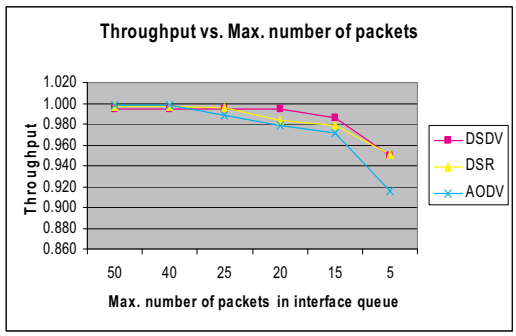


Figure 4: Throughput for each protocol versus different values of maximum packets in the queue.

7. Conclusion

In this paper we have presented the NS simulation results as well as the calculation of Throughput versus three parameters:

1. Speed: by examining Figure 2 we notice that the throughput increases as the node speed increases in the beginning till it reaches a steady state between speed of 5 m/s and 50 m/s. As the speed increases above 50 m/s, there is a decrease in the throughput. Although DSR shows a slight better performance than DSDV and AODV in the steady state case between 5 m/s and 50 m/s, it performs worse for very low speed as well as speeds higher than 50 m/s.

2. TCP Types: In our experimentation we have examined four types of TCP protocols; Newreno, Reno, Sack1, and Vegas. DSDV has performed worse in all the four types of TCP protocol when compared to DSR and AODV. AODV has performed better than DSR except in the case of Vegas.

3. Maximum number of packets: The throughput for the three routing protocols is almost constant for a maximum queue length greater than 25. As the maximum queue length decreases the throughput decreases. For the same queue length DSDV performs better than DSR which performs better than AODV.

References

1. The Network Simulator – ns2. http://www.isi.edu/nsnam/ns/
2. Marc Geris, NS Tutorial, available at: http://www.isi.edu/nsnam/ns/tutorial
3. Yi Lu, Weichao Wang, Yuhui Zhong, Bharat Bhargava, "Study of Distance Vector Routing Protocols for Mobile Ad Hoc Networks," Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003.
4. Rendong Bai and Mukesh Singhal, "DOA: DSR over AODV Routing for Mobile Ad Hoc Networks," IEEE Transactions on Mobile Computing, Vol. 5, No. 10, October 2006, pp. 1403 – 1416.
5. Eitan Altman and Tania Jimenez, NS Simulator Course for Beginners, Chapter 9, pp. 111 – 129. http://www-sop.inria.fr/mistral/personnel/Eitan.Altman/ns.htm
6. Mobile Communications, 2nd Edition by Jochen Schiller, Publisher: Addison Wesley, 2003.

Optimizing Bandwidth Usage and Response Time Using Lightweight Agents on Data Communication Network

E.A. Olajubu, G.A. Aderounmu, and E.R. Adagunodo
emmolajubu@oauife.edu.ng, aaderoun@oauife.edu.ng and eadagun@oauife.edu.ng
Department of Computer Science and Engineering
Obafemi Awolowo University, Ile-Ife. Nigeria

ABSTRACT
In this paper we propose light-weight-agent as an efficient and effective tool for implementing data communication network bandwidth optimization and response time. The routing model for the agents is based on TSP algorithms. Performance comparison was carried out among RPC, Single Mobile Agent and tiny agents otherwise referred to as lightweight agents on the basis of bandwidth usage and response time. The performance evaluation shows the superiority of lightweight agents over the other schemes in terms of bandwidth usage and the response time.

KEY WORDS
RPC, TSP, Lightweight agents, Response time.

1. Introduction

The exponential growth of ubiquitous network has attracted the significant attention of researchers to evolve supportive software that will guarantee some desirable level of QoS in communication network to the end users. This desire has led to proliferation of network-centric supportive applications with little or no attention on how such supportive (network management) software makes use of available bandwidth. Many of these applications software are bandwidth intensive [1][2]. The need for supportive software that minimizes bandwidth usage is very important in today's data enterprise network and more importantly in the developing world where bandwidth is relatively expensive when compared with the advanced economy [20] [3] where the economy is not viable enough to support the scarcity of bandwidth. The advent of network-centric tools has greatly improved researches on mobility of software on communication network. The advantage promised by these technologies (Remote Procedure Call (RPC) and mobile agent) will be little realized without necessary consideration of how efficient they are able to optimize bandwidth usage.

This paper proposes lightweight agents for bandwidth optimization and fast response. The rest of the paper is organized as follows: section two discusses related works while section three explains the routing model. In section four we discuss simulation results and the last section concludes the paper.

2. Related Works

The management of an enterprise network requires that the core applications receive a desirable level of bandwidth attention rather than supportive application software that merely guarantee some level of QoS. The supportive application meant to increase the QoS of the network should minimize bandwidth usage to the nearest minimum. It is often accepted that the higher the bandwidth that is acquired on the network, the higher the throughput of that network without experimental proof. Throwing excessive bandwidth is a trivial solution to guarantee a high QoS expected by the end users without minimization of bandwidth usage by supportive application [10]. It is necessary for supportive applications on the network to economize bandwidth usage thereby leaving substantive portion of the bandwidth for the core applications on the network. Therefore, this paper presents lightweight agents as an alternative tool for bandwidth optimization on data communication network. Agents are self-contained software processes that run in the background and perform a useful task on behalf of a user or group of users. Agents come in variety of models. Simple agents perform small, repetitive tasks, while intelligent agents can decided which tasks are to be included in the processing by choosing from the library of simple agents to perform several tasks in proper sequence [21]. The agents in this work are simple agents called lightweight agents. Lightweight agents are tiny mobile agents that posse limited intelligence in their domain but their communication with each other and interaction with their environment is capable of exhibiting collective intelligent behavior. The agents here are entities capable of sensing their environment and undertaking simple processing of environmental observations in order to perform simple computations. A comprehensive intelligence emerged as a result of these limited intelligent agents interacting locally with one another and especially with their environment to produce an overall intelligence as end result. They are similar to swarm intelligence [14][15] but do not exhibit swarming pattern of movement. The agents are independent of each other in network routing, the routing pattern of each agent depends on its task and present position but all agents use Traveling Salesman Problem (TSP) which is known to be NP hard [22] algorithms to route the network. The agents accomplish their essential tasks with minimal code. They are smaller in size, simpler and faster in action [19] when

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 335–340.

compared with a single mobile agent that has all the intelligence required to perform a task. Mobile agent has been viewed as an omnipotent tool for implementing code mobility for network supportive systems [4][17] due to the fact that mobile agent is capable of suspending execution on one node and migrates to another machine to resume execution from the point where it stopped, while RPC holds on to the network throughout its period of communication.

There have been tremendous research findings in favor of multi-agents usage for data communication network management [5][6][7][8][9][12]. These architectures propose multi-agent systems without due consideration for implications on bandwidth usage. Though mobile agent is a widely acclaimed technology that reduces network loads (minimizes bandwidth usage) through remote computation, yet when the agents are very numerous in single application, the code size of these agents may constitute unnecessary bottleneck for the network during their itinerancy [13][11]. Therefore within the context of multi-agents system there is the need to regulate the number of agents routing the network to allow for optimum bandwidth usage especially for supportive software. Too many agents lead to bandwidth wastages in such applications. Mobile agent unlike RPC that holds on to network resources during its communication has been viewed as an omnipotence tool for implementing code mobility for network management systems [4]. The concept of mobile agent path planning has been investigated in [16] which finds the shortest route between adjacent nodes on the network but allows multiple visits to a node is different from TSP. Also, the argument of "succeed-and-stop" [12] which stops an agent when it has got result does not hold for our concept, in our model the agents must visit every node on the network once for network monitoring.

3. Routing Model: TSP

The cost effective path planning of ant-like agent is very synonymous with the TSP which could be stated as follows: An agent wishes to visit an n distinct nodes on the network and return to the home node. The latency between node k and node k+1 is given as $t^l_{k,k+1}$ and also, $t^l_{k,k+1} = t^l_{k+1,k}$. The agent assignment is to find the sequence of the tour so that the overall distance traveled or cost is minimized. Equation 3.1 is the maximum or the longest route the agent can take to accomplish its task, while equation 3.2 describes the routing model for the agent on the network.

Figure 1 shows hypothetical case for this model. The PC_1 is the source node for the scheme. The sub tours that makeup this hypothetical network is shown in Table 1.

Among all the sub tours, the equation 3.2 is to select the optimum path using [18] that minimizes bandwidth usage which also enhances response time. Thus, the most optimum path for lightweight agents in this simple network is the path that minimizes bandwidth usage and at the same time enhances response time. This is the path that follows the thick lines which is sub tour T_1. Generally to specify the order in which the agents visit the nodes is of the form i_1, i_2,i_3…….i_N where N is the last node on the network. The direction of movement of the agents are not important since latency between node k and k+1 is given as $t^l_{k,k+1}$ and $t^l_{k,k+1} = t^l_{k+1,k.}$

Table 1: Sub tours with their corresponding latencies.

Tours	Latency between two adjacent nodes			
T1	PC_1PC_2	PC_2PC_4	PC_4PC_3	PC_3PC_1;
T2	PC_1PC_3	PC_3PC_2	PC_2PC_4	PC_4PC_1;
T3	PC_1PC_2	PC_2PC_3	PC_3PC_4	PC_4PC_1;
T4	PC_1PC_4	PC_4PC_2	PC_2PC_3	PC_3PC_1

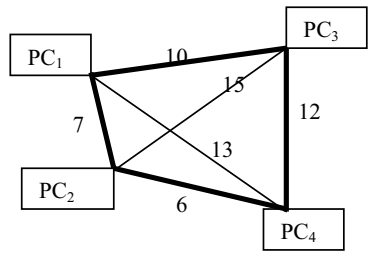


Figure 1: Simple Routing Scenario

In our hypothetical network above, the bandwidth cost between node k and k+1 for each sub tour is as presented in Table 2.

Table 2: Bandwidth cost

Sub tours	*Associated bandwidth Cost*				*Total*
T_1	7	6	12	10	35
T_2	10	15	6	13	44
T_3	7	15	12	13	47
T_4	13	6	15	10	44

Shown in equation 3.1 is the expected longest route on the network. The longest route can be defined to have the highest bandwidth usage and also have the longest response time. Therefore, Φ_{max} in reference to bandwidth usage is the route that consumes the highest bandwidth. Also, in terms of response time, it is the route that gives the longest response time [11].

$$\Phi_{\max} = Max(T(n_i)), 0 \le i \le N \tag{3.1}$$

The general routing model for mobile code on the network is given as

$$Minimize Z = \sum_{k=1}^{N} \left(t^l_{k,k+1} + t^c_k\right) X_{k,k+1} \le \Phi_{\max} \tag{3.2}$$

Table 3: Summarizes the used notations in this paper

Symbols	Description
Φ_{max}	Maximum latency for the longest tour

$t^l_{k,k+1}$	Latency between node k and node k+1
t^c_k	Computational time on the node k
n_i	Node identifier
S	Home node
N	Total number of nodes on the network
X	Indicator of sub tours which are included in the agent path.
T	Tour
Z	Optimal parameter to be minimized
β	Bandwidth usage
X_{ag}	Agent code size in bytes
X_{lwa}	Lightweight agents code size in bytes
y or z	message and response size in bytes
A_i	Number of agents involved in network routing
A	number of communication links that must be established

The purpose of this section is to develop mathematical model for bandwidth β usage against the size of the network i.e. number of nodes on the network for LWA, RPC and SMA. One of the essential commodities on communication network is bandwidth, thus as network availability increases the demands for bandwidth also increases; therefore the need to optimize bandwidth usage especially among supportive software on data communication network is a necessary task which should be done. The bandwidth consumption rate of the three schemes will therefore be simulated to select the one with low bandwidth usage.

3.1 Bandwidth Usage

3.1.1 Single Mobile Agent (SMA) Bandwidth usage equations

In mobile agent scenario, the agent arises from the home node (which acts as the server for the network) and visits all the nodes in the network as single batch and the result are sent to the home node. During visitation time the message size is assume to zero (i.e. y = 0) since the agent only migrate to the nodes with its code size X_{ag} and agent must return with the same code size, therefore the size for the complete journey $2X_{ag}$ and number of nodes in the network is N. But the reply size is represented as z bytes then the total bandwidth usage β for mobile agent is given by

$$\beta_{sma} = \sum_{j=1}^{N} (2X_{ag} + z) \tag{3.3}$$

3.1.2 Lightweight Agent (LWA) Bandwidth Usage:

As in the case of single mobile agent, the agents visit the nodes with their code size X_{lwa} and also return with X_{lwa} code size. The message size y = 0 and the reply size from the nodes z bytes. The number of nodes on the network is N, then the total bandwidth usage in bytes is given as

$$\beta_{lwa} = \sum_{j=1}^{N} \sum_{i=1}^{k} (2X_{lwa} + z) \tag{3.4}$$

Where k is number of agents participating in network monitor

3.1.3 RPC Bandwidth usage

Let the message size from the client be denoted by y and the size of reply from the server be z in bytes, for a single node, the total bandwidth usage in bytes is given by

$$\beta_{rpc} = (y + Z) \tag{3.5}$$

Therefore, for a network that has N nodes

$$\beta_{rpc} = \sum_{j=1}^{N} (y+z) \tag{3.6}$$

If we assume that the message size y (bytes) is the same with reply z (bytes) from the server i.e. y = z and it should noted that a number of links must be established between the two communicating nodes before computation can be done.

Then equation (3.6) becomes

$$\beta_{rpc} = 2\sum_{i=1}^{N} y \tag{3.7}$$

3.2 Response Time

In this research work, the researcher defines response time as the time interval between the time the system sense a faulty/malfunction component and the time the network engineer receives the malfunctioning report. This system is to provide a real-time prediction of fault and notify the network engineer before it really happened. Therefore, if this objective is going to be achieved, the response time of the system must be very high. Therefore, the response time of the three systems will be compared through simulation.

3.2.1 Response Time for SMA

The mobile agent response time is viewed as the time it takes the agent to visit successful nodes on the network. The latency $t^l_{k,k-1}$ is the time taken from node k to node k-1 while t^c_k is the computational time on node k. N is the number of nodes on the network which the agent must transverse. The X is an artificial variable that indicates whether the path is in the present tour or not.

$$MA^{t}_{res} = \sum_{k=1}^{N} \left(t^{l}_{k-1,k} + t^{c}_{k} \right) \quad (3.8)$$

3.2.2 Response time for LWA:

Multi-agent case involved multiple agents cooperating together to monitor various components of network elements. The information is used to determine the present condition of network elements. To accurately determine the response time, the parameter A_i is the number of agents in the same group scanning for the same type of fault on the network element component.

$$LWA^{t}_{res} = \left(\sum_{K=1}^{N} (t^{l}_{k-1,k} + t^{c}_{k}) \right) / A_i \quad (3.9)$$

3.2.3 Response Time for RPC

In RPC scenario, A is the number of requests the source node or home node (which act, as the client) must establish with the node k (which also act as the server) before computation can be done. The other parameters are as found above.

$$RPC^{t}_{res} = A * \left(\sum_{K=1}^{N} (t^{l}_{k,k+1} + t^{c}_{k}) \right) \quad (3.10)$$

4. Simulation Result

Model formulation is carried out for network-centric applications i.e. Remote procedure called (RPC), Mobile Agent (MA) and LWA that can also be used for this supportive software on communication network. MATLAB 7.0 is used to simulate the model formulated.

The agents design is aimed at flexibility, extensibility, simplicity and minimalism. The total code size of an agent is about 15.3 KB, has 9 classes (3public and 6internal) and about 230 lines of codes. The agent code fragment is included in this section.

4.1 Agent Code fragments

```
public class CountAgent extends Agent {
   public static long COUNT_DELAY = 120000;  // thirty
seconds
   public static String countMessageContent = "New
Count is done";
   @Override
   protected void setup() {
   AppUtil.countedAIDs = new Vector<AID>();
  AppUtil.counterAID = getAID();
  addBehaviour(new TickerBehaviour(this,
COUNT_DELAY) {
  @Override
  protected void onTick() {
  AppUtil.countedAIDs = new Vector<AID>();
  try {
     DFAgentDescription template = new
DFAgentDescription();
     ServiceDescription templateSd = new
ServiceDescription();
     templateSd.setType(AppUtil.localServiceType);
     template.addServices(templateSd);

   SearchConstraints sc = new SearchConstraints();
   sc.setMaxResults(new Long(10));

  DFAgentDescription[] results =
DFService.search(CountAgent.this, template, sc);
  if (results.length > 0) {
     for (int i = 0; i < results.length; ++i) {
        DFAgentDescription dfd = results[i];
       AID provider = dfd.getName();
       AppUtil.countedAIDs .add(provider);
       Iterator it = dfd.getAllServices();
       while (it.hasNext()) {
          ServiceDescription sd = (ServiceDescription)
it.next();
         }                }
     catch (FIPAException fe) {
          fe.printStackTrace();

     ACLMessage msg = new
ACLMessage(ACLMessage.CFP);

     msg.addReceiver(AppUtil.controlAID);
     msg.setContent(countMessageContent);
     myAgent.send(msg);
   }
});
   MessageTemplate mt =
MessageTemplate.MatchPerformative(ACLMessage.CFP
);
  ACLMessage msg = myAgent.receive(mt);
                if (msg != null) {
...
```

Performance comparison of bandwidth usage is performed among the three network-centric applications. The bandwidth usage is used as a performance metric for this simulation. Figure 1 shows the degree of bandwidth usage among the three schemes considered. It is obvious in our simulation that tiny agents referred to as light-weight-agents have a better performance than the other two schemes. Though, the bandwidth usage increases with the number of nodes on the networks, the increase in RPC is almost four times more than that of LWA, while single mobile agent scheme is significantly higher than the LWA, making LWA an elegant scheme for implementing network supportive applications that minimize bandwidth usage to the barest minimum.

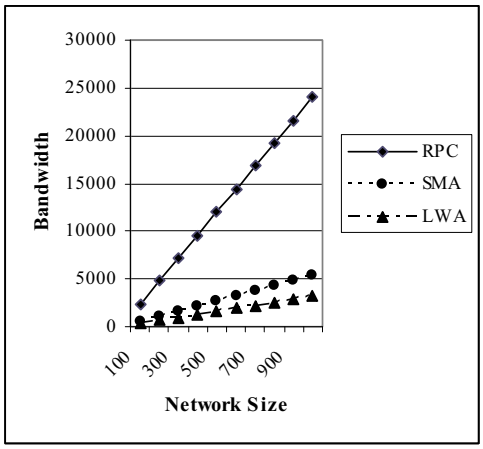


Figure 2: Bandwidth Consumption

The other performance measured for the simulation is response time. Response time in network management is very important as early notification of fault could lead to taking appropriate action that can avert or minimize the effect of the fault(s) on the performance of the network generally. In this simulation the response time of the schemes were also compared. According to our simulation result, any supportive application software based on LWA will have an impressive response time. Figure 3 present the simulation result on response time of the three schemes.

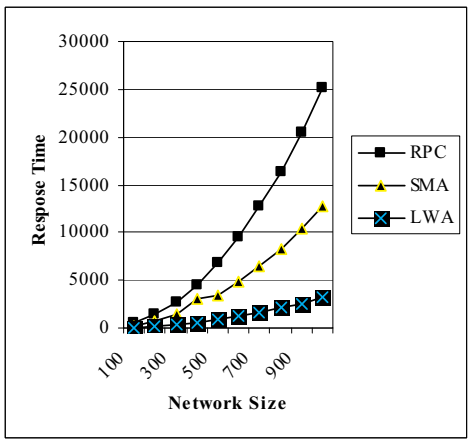


Figure 3: Response time

5. Conclusion

In this paper, we have done a comparative analysis of using mobile codes technology; and the routing model is based on TSP concept. The comparison is based on simulation which is not a real life application. Further simulation work is required for full performance analysis in terms of system throughput, overhead due to retransmission, and system latency. Also, further work is required on information filtering and accuracy of information retrieved for a particular query. Display of information is not precisely organized as will be required by the end user.

References

[1] K. Qureshi and H. Rashid : A performance Evaluation of RPC, JAVA RMI, MPI and PVM . *Malaysian Journal of Computer Science vol. 18 (2)*: 38-44, 2005.

[2] R.J. SouZa and E.E Balkovich Impact of hardware interconnection structures on the performance of decentralized software Proceedings of the 8th annual symposium on Computer Architecture pp. 357 - 365 , 1981.

[3] **M.** Blocksome , C. Archer, T. Inglett, P. McCarthy, M. Mundy, J. Ratterman, A. Sidelink, B. Smith, G. Almasi, J. Castano, G. Lieber, J. MOreira, S. Krishnamoorthy, V. Tipparaju, and J. Nieplocha Blue Gene system software---Design and implementation of a one-sided communication interface for the IBM eServer Blue Gene® supercomputer Proceedings of the 2006 ACM/IEEE conference on Supercomputing pp. 120-135, 2006.

[4] A. Bieszczad, B. Pagurek, and T. White Mobile Agents for Network Management IEEE Communication Surveys. Available at www.comsoc.org/pubs/surveys [Aug. 10, 2006]

[5] K. Boudaoud andZ. Guessoum A Multi-Agents System for Network Security Management. Proceedings of the IFIP TOS WG6.7 Sixth International Conference on Intelligence in Networks: Telecommunication Network Intelligence pp. 407-418, 2000.

[6] A.S. Torrellas Gustavo and A.V. Vargas Luis Modeling a flexible Network Security Systems Using Multi-Agent Systems: Security Assessment Considerations. Proceedings of the 1st International Symposium on Information and communication technologies pp. 365-371, 2003.

[7] M. F. De Castro, H. Lecarpenttie, L. Merghem and D. Gaiti An Intelligent Network Simulation Platform Embedded with Multi-Agents Systems for Next Generation *Internet. Telecommunications and Networking-ICT* pp. 1317-1326, 2004.

[8] N. Rouhana and E. Horlait Dynamic Congestion Avoidance Using Multi-Agents Systems . Proceedings of the third International Workshop (NATA 2001) on Mbile Agents for Telecommunication Applications. Pp. 1-10, 2001.

[9] L. Won-Jong, K. Hyung-Rae, P. Woo-Chan, K. Jung-Woo, H. Tack-Don and Y. Sung-Bong A New Bandwidth Reduction Method for Distributed Rendering Systems Proceedings of First EurAsian Conference on Information and Communication Technology pp. 387-394, 2002.

[10] A.O Oluwatope, G.A. Aderounmu, E.R. Adagunodo and A.D. Akinde Stochastic Reward Net End-to-End Quality of Service (QoS) Simulation Modeling across ATM Network : Using the Goodput Model Proceeding of the IEE Telecommunication and Quality of Service: The Bussiness of Success (QoS 2004) pp.165-170, 2004.

[11] Baek, J., Kim, J., and Yeom, H.. Timed Mobile Agent Planning for Distributed Information Retrieval Proceedings of the fifth international conference on Autonomous agents pp **1**20 - 121 , 2001.
[12] Baek, J., Kim, G., Yeo, J., and Yeom, H.. Cost Effective Mobile Agent Planning for Distributed Information Retrieval Proceedings of the The 21st International Conference on Distributed Computing Systems pp. 65-72, 2001.
[13] Boutaba R., Iraqi Y., and Mehaoua A. A Multi-Agent Architecture for QoS Management in Multimedia Networks. *Journal of Network and System Management Vol.11* (1): 83-107, 2003.
[14] Jennifer Kay and Julius Etzl and Goutham Rao and Jon Thies The (ATL) Postmaster: A System for Agent Collaboration and Information Dissemination Proceedings of the 2nd International Conference on Autonomous Agents (Agents'98), pp. 338--342", 1998.
[15] T. White.and B. Pagurek., Towards Multi-Swarm Problem Solving in Networks, Proceedings Third International Conference on Multi-Agent Systems (ICMAS '98), pp. 333-340, 1998.
[16] K. Moizumi and G. Cybenko The Traveling Agent Problem *Mathematics of Control, Signals and System*, 1998.
[17] B. Brewington, R. Gray, K. Moizumi, D. Kotz, G. Cybenko, and D. Rus Mobile Agents in Distributed Information Retrieval , *In Intelligence Information Agents* pp. 355-395, 1999.
[18] C. W. Duin Two fast algorithms for all-pairs shortest paths. Journal of Computers and Operations Research pp. 2824-2839, 2007.
[19] Helmer G., Wong J. S. K, Honavar V., Miller L., and Wang Y. Lightweight agents for intrusion detection *The Journal of systems and software 67* pp. 109-122, 2003.
[20] The International Network for the Availability of Scientific Publications (INASP) Optimizing Internet Bandwidth in Developing Country Higher Education" A study presented at the Regional Training Conference on Improving Tertiary Education in Sub-Saharan Africa*:* Things That Work*!* Accra, September 23-25, 2003 at http://www.inasp.info/pubs/bandwidth/
[21] D. Seo and B. Moon Computing the epistasis variance of large-scale traveling salesman problems . In Proceedings of the 2005 Conference on Genetic and Evolutionary Computation. Pp.1169-1176, 2005.
[22] R. Tanler. The Intranet Data Warehouse John Wiley & Sons, Inc 1997.

Location Information Discovery for IP Telephony

Leon Stringer Frank Dudek Carolin Bauer

Staffordshire University
Beaconside, Stafford, ST18 0DG
United Kingdom
f.dudek@staffs.ac.uk

Abstract - **The lack of an accepted standard approach to the provision of caller location information is a serious shortcoming in IP telephony. This paper presents a novel approach to implementing this functionality without the additional infrastructure overheads typically found in comparable efforts. The proposed system achieves caller location identification using a series of SNMP requests that could be implemented on existing networks. Initial evaluation shows potential for further development as an alternative to other proposed solutions.**

I. Introduction

The lack of a standard mechanism for identifying caller location within IP telephony systems ("VoIP") assumed a new level of importance when its increasing adoption across market sectors started to attract the attention of various regulatory bodies such as the FCC in the USA [7] and Ofcom in UK [13]. Within the traditional Public Switched Telephone Network (PSTN), approximate location information can be determined for a fixed line telephone from the number's area code, and network operators may make more precise information available to authorised users for law enforcement and emergency use. IP-based computer networks lack equivalent mechanisms for use with IP telephony, problematic not only in terms of consumer expectation, but increasingly in complying with statutory frameworks for telecommunications operators and public safety requirements.

Ongoing research has led to increasingly sophisticated techniques for determining the location of Internet hosts, but these are too general for the IP telephony applications described. Work is also continuing to develop services ranging from proprietary techniques for individual systems to the introduction of new open standards for wider adoption. Such services typically involve an additional tier of functionality to existing network systems which could be viewed as a complicating factor impeding adoption.

These approaches along with other contextual factors are presented in the following section. This is followed by a description of a new mechanism for providing location information sufficient for use with IP telephony applications, notably emergency calls. This approach is based on existing protocols and equipment as the motivation for this research is to produce a solution with a minimum of prerequisites, thus lowering the barrier to adoption. We will show that only the telephony client firmware requires some changes with location information storage distributed throughout the network. The implementation of a prototype to further evaluate the feasibility of the approach is then described and finally a conclusion summarises the outcome of this work

II. Background

IP networks are location transparent: addresses for individual computers bear no relation to the physical location of the device and network device location may change without any modification to the address. So while IP-based telephone systems are more flexible in terms of deployment than non-IP systems, locating callers on IP telephony systems is challenging. Besides the possibility of enabling innovative location-based services, automatically establishing caller location is a cornerstone of the functionality expected for emergency calls [7, 13]. This is for two reasons: to direct the call to the correct emergency authority for the caller's region and then expedite the dispatching of aid to the correct location. The first requirement only requires approximate location information whereas the second requires more accurate details such as a street address.

Techniques for remotely determining the location of Internet hosts ("geolocation") are insufficiently accurate and too error prone for this kind of application. Published techniques such as IP2Geo [14], constraint based geolocation [9] and their proprietary counterparts have proved sufficient for some less demanding applications (e.g. targeted banners for web advertising). However, such techniques are typically unable to traverse network gateway services, e.g. identifying the location of a proxy server when the actual host is potentially a great distance from this location.

GPOS [6] and LOC [3] are two DNS-based services which offer a way to store location information for hosts. The latter is the most well-known but has seen negligible uptake primarily because there is little motivation to maintain such information. Another obstacle may be its reliance on longitude and latitude to describe location, information not readily available to most people.

Consequently, IP telephony-specific solutions emerged involving some form of managed infrastructure service on the local network where user agents (e.g. IP phone or desktop application) or call agents (e.g. IP-capable PBX or SIP proxy)

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 341–345.

can obtain location information for onward transmission. Systems designed for use with existing emergency call centres typically provide a telephone number appropriate for that location. In this way, the call centre establishes the location for the telephone number as it would for any other call [2]. However, with the increasing use of IP telephony systems, work on envisaging the operation of an IP-only emergency call service with location information is supplied directly – as a street address or using some form of coordinate system – is under way.

Manually entering location information via the user agent (UA) avoids a great deal of complexity. The consumer VoIP service operated by Vonage stores user-submitted location information in a central location information server (LIS) for use by their operators in the event of an emergency call [18]. Although workable, this seems like a retrograde step compared to PSTN services and failure to keep location information up-to-date could have serious consequences when attempting to summon help. However, manual input may be appropriate when other mechanisms to identify location have failed.

Mintz-Habib, Rawat, Schulzrinne & Wu [12] propose the use of location aware wireless devices accessible by UAs. The use of such devices, accessed by Bluetooth, Wi-Fi, RFID, etc., is similarly proposed in [8] and also in [19]. However, the desirability of this approach is questionable. UAs would have to support a variety of wireless technologies and be within range of at least one location-aware device. The devices themselves would need some standard way to serve location information as well as somehow becoming initially aware of their own location. Considerable development in the use of wireless location-aware devices is required to make this service generally available and where it is not, Mintz- Habib, et al. propose falling back to a database mapping MAC addresses to locations.

For location by MAC address, a database could be created and maintained by system administrators. UAs could look up their own location or call agents could automatically determine and include the information in the call where UAs do not support this functionality [12]. The UA's MAC address is a logical choice for the database key but this information may not be available to a call agent on a different subnet. The IP address is an obvious alternative but requires the call agent to be part of the same IP addressing scheme (i.e. with no NATing, etc.).

Another potential key could be the combination of network switch and switch port. Kim, Song & Schulzrinne [10] propose the use of Cisco Discovery Protocol (CDP) to achieve this. UAs listen for CDP advertisements then query the LIS for the location information for their unique port. A Telecommunications Industry Association working group [8] are proposing a similar system using CDP's recently published open standard equivalent LLDP-MED [17]. LLDP-MED also offers the possibility of directly advertising location information in coordinate, street address or telephone number form. For each individual port, a suitably capable UA can listen for this information and either cache it locally or register it with an LIS.

Where LLDP-MED is not available, [8] describes how an LIS may interact with a call agent to determine which devices are IP telephony capable and then discover the location of these by examining network switch forwarding tables with SNMP, then querying the LIS database for the relevant port location. Changes to these connections are tracked using SNMP traps. Potential problems again arise over how the call agent identifies the devices as the switches will store MAC addresses only.

Clearly this is an area of ongoing work. The ability of LLDP-MED to directly advertise location information to connected devices is likely to become the standard mechanism for IP telephony applications in the future. The only drawback is its relative nascency with many existing devices not capable of this interaction. By focusing on the core problem of discovering location information and how this might be achieved with existing technologies the next section will describe an alternative approach with comparable features but less reliance on complex infrastructure elements.

III. IPTELLIDISCO

A. Overview

Clearly new infrastructure services are needed to obtain the accuracy required by IP telephony applications, but proposed and emerging systems involve a significant overhead in terms of additional equipment and systems to install and maintain. With the aim of reducing this overhead and thus the obstacles to adoption, the following approach was considered firstly as a theoretical framework and then prototyped to further evaluate its feasibility. The cornerstone of this approach, which we call IPTelLIDisco, is for a user agent to determine the switch to which it is connected and obtain location information from this. This is achieved by initially querying all devices in the UA's subnet using SNMP.

The Simple Network Management Protocol [15] and its associated standards provide a mechanism to get and set status and management information of network devices and provide notifications of alert states. SNMP's architecture separates the protocol from the information that can be managed. This information is defined in Management Information Bases (MIBs) which include a core set of common values defined in MIB-II [16], and sets of information for specific classes of devices. Managed network switches support the Bridge MIB [5] which allows access to the switch's forwarding table, i.e. a list of MAC addresses the switch knows about and the ports these are connected to. Switches may be interconnected ("cascaded") so there may be multiple MAC addresses listed with the same port value where this port connects to another switch.

Thus SNMP enables a client to query a switch and search for its MAC address and thus determine if it is connected to that switch. MIB-II includes the ability to assign a location description to a device intended for network management e.g., 'telephone closet, 3rd floor'. This field – sysLocation – at 255 characters could also include a street address and each switch can have a different value. The discovery of the location information from a managed switch is shown in Figure 1.

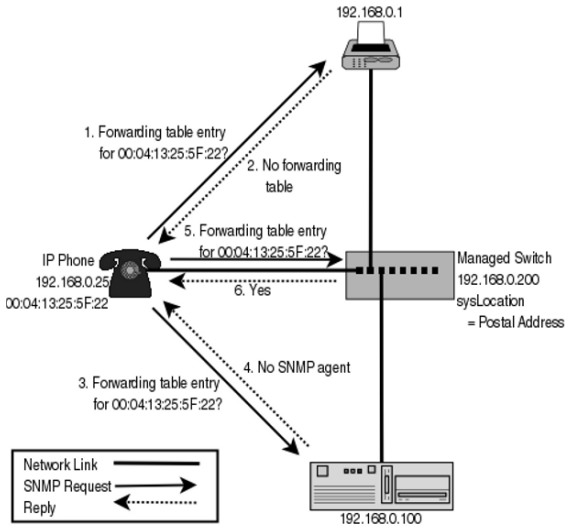


Figure 1. Discovering location information from the connected switch

Whilst this method is less accurate than the port-level resolution offered by systems such as [10], it is potentially more accurate than that available with PSTN systems and so presumably sufficiently accurate for telephony applications including emergency calls. Location information, once discovered, could be cached for long periods before being refreshed while the user agent remained connected to the network. When a call requiring location information is made, the information can be sent as part of the call set-up information, the approach taken by [12].

B. Prototype

An implementation of this idea in prototype form was undertaken to further evaluate its feasibility. Rather than develop a complete user agent, the open source softphone Ekiga (http://www.ekiga.org) was used since it could be altered and extended as required. The search for the connected switch is undertaken by iterating through the subnet address space querying each potential device. Although this may seem time consuming, this process would occur in the background when the client was activated and stored locally ready for use when an appropriate call was made.

The prototype uses SNMPv2 and assumes the switch uses the default read-only community string “public”. The process of location information discovery requires the user agent to find its MAC address in the switch’s forwarding table with no other MAC addresses with the same port value. Ordinarily one must iterate through the forwarding table as the indexes to each MIB entry in the forwarding table are unknown. But the index values turn out to be the individual bytes of each MAC address. As these are known, the user agent can directly query the forwarding table using its local MAC address value. Not only does this offer a short cut to determine if a switch has a forwarding entry for the client but, by sending this query to each device in the discovery process, rapid determination of the following can be made in a single exchange of packets: whether there is a host at that address, whether it is SNMP-capable, whether the host is a switch and, if so, whether it has a forwarding entry for the client. This process requires just one packet per host (plus any response).

When a connected switch is found, an iteration of forwarding table is required to determine if the connection is direct. Indirect connections are remembered and rated by their topological proximity to the UA as these can be used if no direct connection is found. Before doing this however, the location information is queried to determine if this is useful before bothering with the expense of iterating through the forwarding table.

The location information stored in the sysLocation field is described as a street address. In order to improve the consistency of the format for storage and processing, address fields are formatted according to the rules in the vCard specification [4] and separated from any network specific location information (i.e. that for which sysLocation was intended) by an additional delimiter. In this way the client can quickly determine if a switch has usable location information by checking for the vCard fields, making it easily computer and human-readable without the verbosity of an XML-based format such as that proposed by [19].

Location information from an indirectly connected switch can be used if no direct connection is found. When the discovery process is complete, a notification message appears reporting the location so the user can report any suspected errors. As a last resort, if no location information is discovered, a dialog window allows the user to enter their address manually.

The obtained location information is cached between sessions along with a value indicating when this information is considered out-of-date and beyond which the auto-discovery process is recommenced. Where the location is determined to have been obtained from a directly connected switch, the IP address of this switch will also have been stored and is then queried first. With indirect connections the discovery process occurs in full, otherwise the system would fail to discover a new direct connection introduced after the previous discovery.

In the prototype the UA is preconfigured with SIP URLs for which location information is required. When a call is placed to one of these, location information is sent as text content to the recipient alongside the call. Before this is sent, user permission is requested. However, mindful of the possibility of its use in an emergency, if the user takes no action the information is sent automatically after a timeout period. It is envisaged that for emergency calls a set of well-known URLs will emerge for which the UA would automatically send location information.

The location information is sent to the called party using the SIP MESSAGE extension [1]. This was designed to support text messaging applications such as instant messaging (IM) and coexists independently with any voice calls. In the

prototype the information is sent when the call connects although for some applications it may be more appropriate to send this as soon as the call setup attempt is made.

C. Analysis

The prototype implementation demonstrated that this is a potentially feasible way to provide location information to UAs for use in IP telephony applications. Location information is retrieved on the LAN and sent in parallel with the voice call, and there is no problem traversing proxy servers or similar potentially problematic services – if the call can be established, the location can be sent.

Distributing the storage of location information into the switches rather than a central LIS makes the approach highly scalable. However, operation of this system is based on two premises: that the switch's SNMP interface is on the same subnet as the client, and that the SNMP community string is known. Although the first issue is likely to be mainly a problem for VLANs, there's no requirement for switches to occupy the same subnet as clients even when they occupy the same physical LAN. The SNMP queries require a read-access "community string" (a plain-text password value). The easiest solution would be for all switches to use the common default "public". However, this could be regarded as introducing unnecessary risk to network security. Alternatively UAs could be configured with a community string, based on the reasonable expectation that an organisation uses a single set of community strings across their network. A further improvement would be to establish a common standard string with restricted access. SNMPv3 overhauls SNMP authentication replacing community string-based authentication with passworded individual accounts with specific levels of access. This would allow configuration of a well known user account with read-only access and limited to the sysLocation and switch forwarding table MIB objects.

Another potential weakness is the discovery process itself which could be considered primitive. However, since there are other applications for this kind of technique (e.g. network management tools) there has been work undertaken to develop more sophisticated network discovery techniques [11]. Where UAs are directly connected to switches, the discovery process need only run once per user agent.

Overloading sysLocation may cause problems for some network operators who use this field as intended although merely adding more location information is not an unreasonable use of this field. This situation could be improved by introducing new MIB objects specifically designed for this purpose. Such objects could include individual civic location fields (e.g. street, city, postal code) and fields for coordinate systems (e.g. longitude and latitude). LLDP-MED includes an optional MIB for the new functionality including location information. However, this is stored in an SNMP table on a port-by-port basis rather than a single switch-wide object. Standardising authentication and access controls for access to a set of objects for geolocation is presumably similar to the goals of those working on geolocation via wireless beacons.

As with other solutions based on SNMP, this system fails to cater for home users. Home IP telephony users' network connectivity is typically via some form of broadband router, often incorporating a Wi-Fi access point (AP). Although SNMP support on such devices is increasingly common, even in entry-level devices, this is usually restricted to MIB-II objects only. However, as high-end APs (e.g. the HP ProCurve 520WL) support the SNMP Bridge MIB, presumably if there was demand, such functionality could be easily included in consumer oriented devices.

Transferring location information as plain text may be considered an unacceptable security risk. This could be alleviated by some form of public key cryptography system. The location information could be stored as cypher text on the switch and consequently is never transferred in unencrypted form. An asymmetric system generating public and private key pairs for both parties would prevent location details inadvertently sent to unintended parties from being read, as well as giving the recipient some guarantee of the sender's identity. Despite these issues, the system is capable of doing what it set out to do: provide sufficiently accurate location information to telephony clients through an automatic process. Taken in isolation, it appears to have significant drawbacks, as outlined above. However, in the context of other systems, the problems do not significantly detract from its feasibility.

IV. Conclusion

A new approach to automatic location information discovery using managed switches has been described. This solution has been presented in the context of existing and proposed alternative methods. In testing this has been found to be a feasible approach to provide this service on a small network. Although this approach does have some weaknesses these could be mitigated through further work such as the use of encryption to provide security and improvements to the discovery process to improve scaling of this method.

The approach can be significantly enhanced through the introduction of MIB objects for network switches specifically aimed at location-aware applications. Combined with a standard SNMPv3 account allowing access for any device to this information, this could create a common platform for the provision of location information. For existing switches, this functionality could be provided by manufacturers in a firmware update.

REFERENCES

[1] Campbell, B., J. Rosenberg, H. Schulzrinne, C. Huitema & D. Gurle, 2002, RFC 3428 – Session Initiation Protocol (SIP) Extension for Instant Messaging, retrieved from the Internet: ftp://ftp.rfc-editor.org/in-notes/rfc3428.txt, January 05, 2007

[2] Cisco, 2004, Cisco Emergency Responder Version 1.2 (2), retrieved from the Internet: http://www.cisco.com/application/pdf/en/us/guest/products/ps842/c1650/ccmigration_09186a008009253a.pdf, August 24, 2006

[3] Davis, C., P. Vixie, T. Goodwin & I. Dickinson, 1996, RFC 1876 – A means for expressing location information in the domain name system, retrieved from the Internet: ftp://ftp.rfceditor.org/innotes/rfc1876.txt, July 20, 2006

[4] Dawson, F. & T. Howes, 1998, RFC 2426 – vCard MIME Directory Profile, retrieved from the Internet: ftp://ftp.rfceditor.org/in-notes/rfc2426.txt, January 11, 2007

[5] Decker, E., P. Langill, A. Rijsinghani & K. McCloghrie 1993, RFC 1493 – Definitions of Managed Objects for Bridges, retrieved from the Internet: ftp://ftp.rfceditor.org/innotes/rfc1493.txt, October 17, 2006

[6] Farrell, C., M. Schulze, S. Pleitner & D. Baldoni, 1994, RFC 1712 – DNS Encoding of Geographical Location, retrieved from the Internet: ftp://ftp.rfc-editor.org/in-notes/rfc1712.txt, May 14, 2007

[7] FCC, 2005, In the matters of IP-enabled services E911 requirements for IP-enabled service providers: First report and order and notice of proposed rulemaking, retrieved from the Internet, http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-05-116A1.pdf, April 3, 2006

[8] Gass, R. (ed.), 2007, IP Telephony Support for Emergency Calling Service (TSB-146-A) [Unpublished draft], retrieved from the Internet, http://ftp.tiaonline.org/TR-41/TR41.4/Public/2007-02-Houston/TR41.4-07-02-006-TSB146-AFinalWithrevisionmarks. doc, May 04, 2007

[9] Gueye, B., A. Zivani, M. Crovella & S. Fdida, 2004, Constraint-based geolocation of internet hosts, Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, Taormina, Sicily, Italy, ISBN 1-58113-821-0, pp. 288-293

[10] Kim, J.Y., W. Song & H. Schulzrinne, 2006, An enhanced VoIP emergency services prototype, retrieved from the Internet: http://ng911.tamu.edu/portals/ng911/story_docs/NG911_ISCRAM2006.pdf, September 05, 2006

[11] Lowekamp, B., D. O' Hallaron & T. Gross, 2001, Topology discovery for large ethernet networks, Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications, San Diego, California, United States, ISBN:1-58113-411-8, pp 237-248

[12] Mintz-Habib, M., A. Rawat, H. Schulzrinne & X. Wu, 2005, A VoIP emergency services architecture and prototype, Proceedings, 14th International Conference on Computer Communications and Networks, ISSN 1095-2055, pp. 523-528

[13] Ofcom, 2006, Regulation of VoIP Services: Statement and further consultation, retrieved from the Internet, http://www.ofcom.org.uk/consult/condocs/voipregulation/voipregulation.pdf, April 3, 2006

[14] Padmanabhan, V.N. & L. Subramanian, 2001, An investigation of geographic mapping techniques for Internet hosts, Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications, ACM Press, New York, ISBN 1-58113-411-8, pp. 173-185

[15] Presuhn, R. (ed.), 2002, RFC 3416 – Version 2 of the Protocol Operations for the Simple Network Management Protocol (SNMP), retrieved from the Internet: ftp://ftp.rfceditor.org/innotes/rfc3416.txt, January 11, 2007

[16] Presuhn, R. (ed.), 2002, RFC 3418 – Management Information Base (MIB) for the Simple Network Management Protocol (SNMP), retrieved from the Internet: ftp://ftp.rfceditor.org/innotes/rfc3418.txt, January 13, 2007

[17] TIA, 2006, Link Layer Discovery Protocol for Media Endpoint Devices, retrieved from the Internet: http://www.tiaonline.org/standards/technology/voip/documents/ANSI-TIA-1057_final_for_publication.pdf, November 02, 2006

[18] Vonage, n.d., Vonage Provides 911, retrieved from the Internet, http://www.vonage.com/features.php?feature=911, April 03, 2006

[19] Zahid, K.K.A., L. Jun, K. Kazaura & M. Matsumoto, 2004, IP network for emergency service, Proceedings of the 3rd international conference on mobile and ubiquitous multimedia, ACM Press, New York, ISBN 1-58113-981-0, pp. 165-170

A Flow Based Traffic Characterization of IP Telephony Protocols

Geneflides Laureno da Silva
Applied Computer Science Department
University of Fortaleza, Brazil
geneflides@unifor.br

Raimir Holanda Filho
Applied Computer Science Department
University of Fortaleza, Brazil
raimir@unifor.br

Abstract- **This article intends to characterize the traffic flows of SIP/RTP and IAX protocols in the environment of IP Telephony as standard of communication in their local networks. Our study allowed us to conclude that the packages of voice have a significant impact on the amount of data injected in LANs, which makes of fundamental importance the development of tools to evaluate them.**

I. INTRODUCTION

The way we communicate currently, has contributed substantially for cultural changes inside of our society. The main responsible actor for such transformations has been the technology.

In the universe of the communications, mainly in the environment of the computer networks and telecommunications systems, protocols as defined in [1], [2], [3], [4] have played basic role in the evolution of the related structures, having allowed integrations, reengineering and designs of new architectures.

The development of open standards, protocols geared to telephony and multimedia, joint with open source architectures, has allowed to the small companies the access to the use of leading-edge platforms, before restricted to only a portion of the enterprise universe, because their expenses values. Such access, however, has incorporated to the networks a huge volume of traffic, which if is not well measured, will cause inconvenience and damages.

Faced with this reality, where the key word "convergence" is the main pillar of support of these routes through which the information transit, know the behavior of traffic in such networks enables us to a critical and analytical vision of current needs and likely improvements required.

Our work has as main objective, to characterize the IP Telephony traffic behavior, inside of the corporative environment, through the analysis of flows, when of the use of SIP/RTP[4], [12] and IAX [6] protocols.

The IP telephony is emerging in recent years, through the use of VoIP technology [7] as a real evolution of scenarios for traditional telephony, which were based on techniques such as PCM [8] and TDM [9] for transport of voice, but not that the scope range of services, applications and convergence provided by the new scenario with the use of integrated networks for data, mobile and internet.

In section II of our work we will carry through a descriptive briefing of the architecture of the SIP/RTP [4], [12] and IAX [6] to be analyzed, aiming for their contextualization of the ones, within the scope of expertise of small corporations. In section III we describe the scenarios of traffic collection, the method of collection and the filtering of the information, for accomplishment of the intended analysis. The characterization outcomes and the respective graphics are presented in section IV. Finally, in Section V will discuss about conclusions, future works and service enhancements to the filtering techniques employed.

II. SIP/RTP and IAX PROTOCOLS

The SIP (Session Initiation Protocol) can be described in brief as a client-server and P2P architecture, where their operation is based on the standards of the Internet and mechanisms such as those used for sending and receiving e-mails and the web browsers operations. As internet based we would to cite the following protocols: [10], [11].

In Fig.1 we can see the components of SIP architecture and a little description about principal actors.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 346–351.

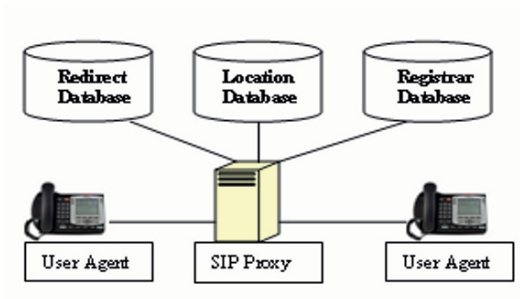


Fig. 1. SIP COMPONENTS ARCHITECTURE

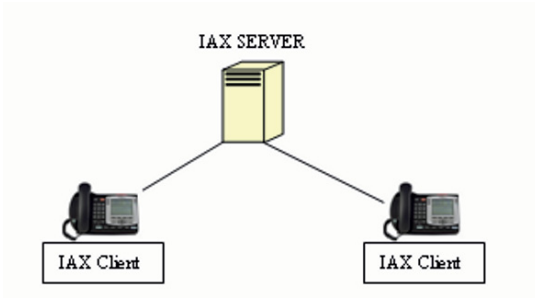


Fig. 2. IAX COMPONENTS ARCHITECTURE

The User Agent is the entity that interacts with the user. It has the ability of send and receive requests. The SIP Proxy is responsible by authenticate, authorization and it acts to send and receive requests to another servers or User Agents. Redirect, Location and Registrar components have the similar functions according respective names.

We will not detail the operation of the SIP, exchanging messages and so on, because more detail can be found in [4]. The main point we would like to report about SIP protocol, is the fact of the use of distinct protocols and *well know ports* of communication for signalling and exchange of packages of voice itself.

In the SIP, after the register of the terminals in the bases of the servers, the exchange of streaming of audio occurs with different numbers of ports, through the performance of protocol RTP [12]. For signalling, the standard port used is UDP 5060 and for the flow of voice, the port can vary between 5000 and 60000. So, there are distinct steps. The first is the signalling with SIP messages to register, authenticate, authorization and call establishment. The second step is the voice packets flow, where the RTP protocol acts. This surely is the primordial factor of differentiation for the IAX [6].

Another important protocol in the performance of the SIP is the SDP [13], which holds the information of the type of media that's being transmitted by RTP.

The SIP protocol has also been highlighted as a basis for new architectures, such as IMS [14], which certainly is emerging as the key structure for FMC (Fixed Mobile Convergence).

The Fig.2 to present components of the IAX architecture, that also works like a traditional Client/Server and peer to peer platform.

We can define IAX as a protocol for control of applications and media, having as functions to create, modify and terminate multimedia sessions on the IP protocol.

The IAX Client exchange requests, responses and voice packets flow. Their main difference with respect to the SIP is accurately the fact of itself carry out the functions of control and dispatch of media, with a single port for action: UDP 4569. Thus does not need a port to make the register and another protocol for sending the conduct of streaming media. This approach simplifies the procedures for administration of firewalls, which require configuration to the rules for just one port. Another important point is to a unification of the signalling and audio pathways, facilitating the transparency of shipping the package, when the use of NAT [15].

The IAX Server is responsible to authenticate, authorization and control of all applications and multimedia sessions.

Although the IAX have more flexibility in certain aspects, the SIP has become the standard used by major manufacturers and issues such as video on phones, integration with small amount of codecs and limitation of use by Softphones and IP phones is restricting the growth of IAX protocol.

The IAX protocol currently is already in version 2 and here is need emphasize its use in conjunction with the open source platform[5], who has won a world increasingly growing in small businesses and a more detailed analysis shows some advantages of their use in such scenarios, as we inquire in our studies and that we will explain in the next sections of our work.

III. TRAFFIC COLLECTING

The scenario of collecting for our traffic samples (traces) were the internal environment of 04 corporations, according to the pattern we can see in Fig. 3. In this topology, several end points(Clients) are connected on the LAN and exchange information with the Voice Server (Server). The customers equipments can be ATAs (Analog Telephone Adapters), IP phones or Softphones.

All these devices uses IAX or SIP protocols to connect and signalling with the Voice Server and are exactly the generators of traffic throughout the network structure, which integrates and connect to the universe of traditional telephony - PSTN(Public Switched Telephone Network).

The new traffic generated within the local network, is the exchange of information between clients and the server. The protocols used for such transactions and the impact of this volume of data in corporate networks become the focus of our research, especially for analysis of the behavior when the use of SIP/RTP and IAX protocols.

In figure 4, we show the steps executed for to characterize the traffic of SIP/RTP and IAX protocols.

We have used the tcpdump [16] software collector, enabling itself in the interface FastEthernet of the Voice Server so that all the traffic that flowed on it, were stored.

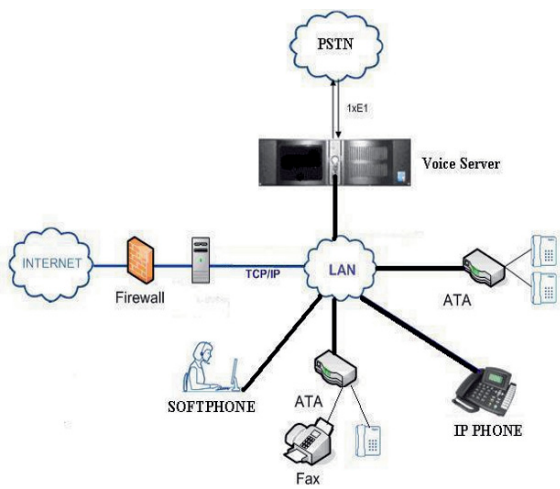


Fig. 3. CORPORATIVE SCENARIO

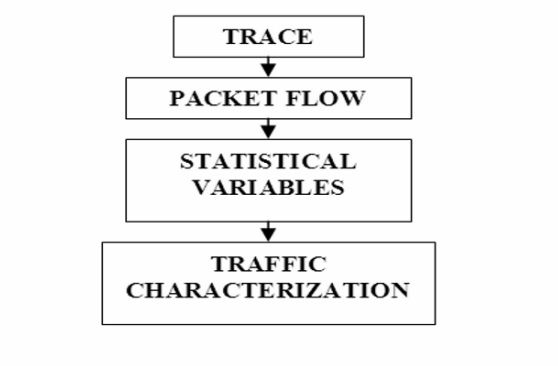


Fig. 4. STEPS TO TRAFFIC CHARACTERIZATION

Each trace collected is referring to a period of 25 minutes, gotten between 10:00 and 10:25 am. In Table I, we show the information about amount of extensions of each corporation, size of files stored and time to gathering. With the files obtained, we used another application called Wireshark [17], in order to export the data from the tcpdump format to txt format and begin the second step.

The second step was to generate packet flow. Based on work [18], we used the definition of a flow as a set of packets that share the following data: source IP address, source port, destination IP address, destination port and protocol.

In step 3 we stand out the importance of the two data information: identification of the amount of packages per flow and the total amount of bytes in each one that we believe in this first study to make possible an aiming in the direction of the important statistical variables that show us a quantitative impact of IP Telephony traffic on local networks.

In step 4 we do the characterization and can evidence aspects of the protocols and characteristics of its flows, what made possible the characterization of traffic.

One of the great difficulties of analysis and characterization of VoIP traffic and by consequence of IP Telephony is the large amount of packages that are generated in the establishment, progress and conclusion of multimedia sessions. We have seen this with this study and the volume of data obtained by a gathering in a space of time so small.

TABLE I

DATA GATHERING

COMPANY	EXTENSIONS	FILE SIZE	TIME
I	46	13 Mbytes	25
II	39	12 Mbytes	25
III	73	145 Mbytes	25
IV	122	424 Mbytes	25

IV. Results

During the signalling processes, both protocols have similar operation, because that the end points will be changing information with the server about locations, passwords and authentications of its accounts.

We detected throughout this stage one strong presence of flows with small amount of packages and consequently a lower percentage of bytes per flow.

In Figure 5, for signalling process, we can evidence the difference of the amount of bytes between SIP and IAX protocols, using flows that have an amount of 1,5,10 and 15 packages, which characterize the process for register, control, establishment and disconnection of a call.

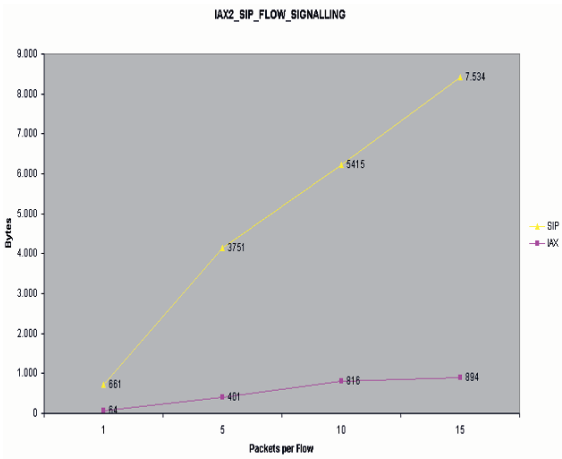


Fig. 5. IAX x SIP SIGNALLING

The interesting one in this analyzes is to perceive that the SIP protocol curve always have values of the amount of bytes per flow higher than that those used by IAX protocol.

In Figure 6 we show a flow size distribution based on the amount of packets per flow. We can see that the SIP protocol shows a behavior where flows with 6 packets are predominant. The IAX protocol, although to present a high number of flows with 5 and 16 packages, is distinguished from SIP, because to present flows with various sizes.

Figure 7 shows us a comparison between IAX and SIP/RTP protocols during calls in progress, after the signalling process. We notice that how much higher the amount of packets in a flow, more steady remains the average of bytes per packet in the flow. In traffic RTP this average revolves around of 214 bytes per packet, while in the IAX this value figure falls to the level of 204 bytes per packet.

In environments where voice calls are more constant and that the volume of bytes per flow surpasses the millions, a difference of 10 bytes per packet will have an impact directly proportional to this number.

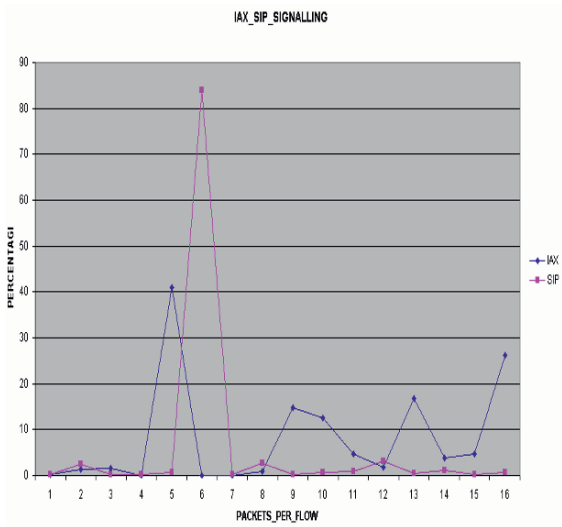


Fig. 6. IAX x SIP SIGNALLING FLOW PERCENTAGE

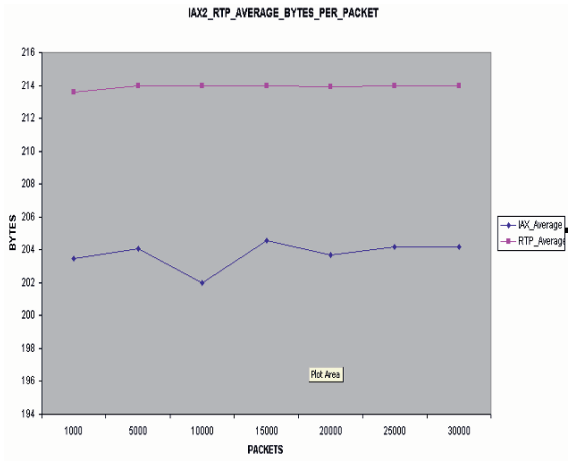


Fig. 7. IAX x SIP/RTP FLOWS: BYTES AVERAGE PER PACKAGE

In Figure 8 is possible to observe in some flow samples, the amount of packets and respective amount of bytes in IAX voice calls. We show samples of flows, so there is no proportional scale in graphic. We intend evidence the volume of packets and bytes in IAX voice flows.

In Figure 9 we show with some flow samples too, the amount of packets and respective amount of bytes in RTP flows. We identify that are voice flows because there are high amount of packets and use of RTP protocol.

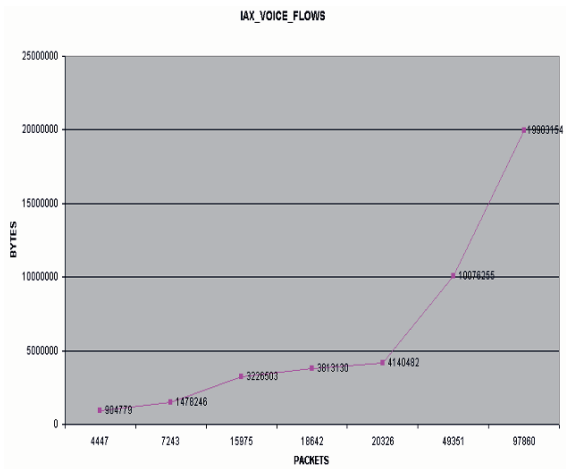


Fig. 8. IAX VOICE FLOW

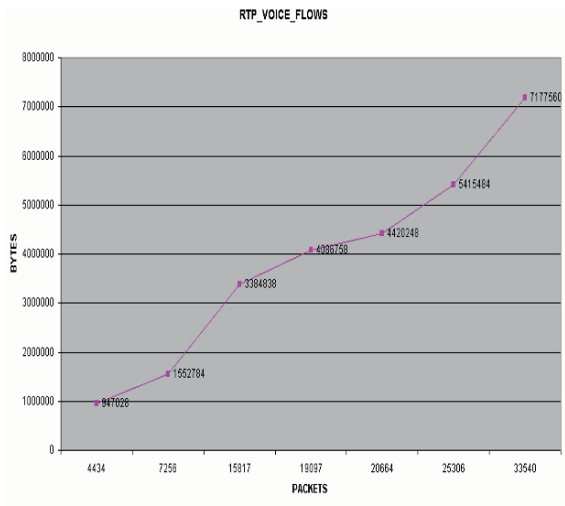


Fig. 9. RTP VOICE FLOW

The Figures 7, 8 and 9 evidence us that the SIP/RTP protocols always have an amount of bytes per flow higher than IAX protocol. If we suppose a voice call with approximately 10000 packages per flow of data, we see that using SIP have always 100K to over information included in the network.

Routers and switches work with queues of mapping packages, so any gain that we may have with economy of the sending of information to accomplish the same task, will be of great value in the scope of work of local area networks.

V. CONCLUSIONS AND FUTURE WORKS

The initial basis of this study allowed us understand the operation of the SIP/RTP and IAX protocols within real corporate environment and show us a important characteristics about the volume and traffic characterization, when of the adoption of IP Telephony in these topologies.

This understanding made us conclude the following considerations about IP Telephony, SIP/RTP and IAX protocols:

1) SIP protocol has always flows with small quantities of packets, and the flows identified as RTP have larger quantities of packages, consequently with high values for the discriminator ‘total of bytes’ per flow.

This behavior ratifies the performance of the SIP, that originally use the port 5060 only for signalling and voice packets follows the RTP flows.

2) IAX protocol follows a pattern with small quantities of packages with small amount of bytes per flow in signalling stage, as well a large quantities of packages with high quantities of bytes per flow in voice stream. This behavior also ratifies the performance of this protocol, which uses the same port for both the functions: control and transport of media.

3) The flows of IAX protocol, having as analysis the two discriminators mentioned previously (total of packages, total of bytes), allow us to evaluate that they always have a higher amount of packages, with a smaller amount of bytes per flow, if compared with SIP flows. This characteristic of proportional inversion also makes us to believe that the impact of the quantitative of bytes inserted ones in a local network, when IAX protocol is used for applications of IP Telephony is lower than when using SIP protocol.

The techniques of VoIP have already spread, have enabled the traffic of voice over data networks and Internet and its use in WAN links has allowed considerable savings for companies with profile connections for distance and international calls. This technology has also allowed the advance of IP telephony in enterprise scenarios, which has contributed to the increase the volume of data passed through in the local networks.

Our study allowed us to conclude that the packages of voice have a significant impact on the amount of data injected in LANs, which makes of fundamental importance the development of tools that can evaluate such traffic, whether focused on the performance of protocols or the mapping of volume of flow of this information. This fact may contribute to a better tracking in the prediction of failure, of intrusion detection and perception of on loads, among others.

We glimpse as future works the following topics:

1)The application of others discriminators based on [18], since the article quoted specifies a sufficiently high number of these possibilities.

2) Explore different combinations of statistical variables.

3) Develop methods to identify and classify the IP Telephony traffic.

REFERENCES

[1]Gary C. Kessler, “ISDN: Concepts, Facilities, and Services”, 1990

[2]Douglas E. Comer, “Internetworking with TCP/IP Vol.1: Principles, Protocols, and Architecture”, 1995

[3] Thom, G.A, “H.323: the multimedia communications standard for local area networks”, Communications Magazine IEEE, Dec 1996

[4] J. Rosenberg, et al., “SIP: Session Initiation Protocol”, RFC 3261, June 2002.

[5] Leif Madsen, Jared Smith, Jim Van Meggelen, “Asterisk: The Future of Telephony”, September 2005

[6] M. Spencer, B. Capouch, E.Guy, F. Miller, K. Shumard, “IAX: Inter-Asterisk eXchange Version 2, IETF draft-guy-iax-01, March 2006

[7] B. Goode, “Voice Over Internet Protocol(VoIP)”, Proceedings of the IEEE, September 2002

[8] KW Cattermole, Principles ot Pulse Code Modulation, London: Iliffe Books, 1969.

[9] WR Bennett, “Time division multiplex systems,” Bell System, Technical Journal, volume 20, April 1941,page 199

[10] J. Klensin, “Simple Mail Transfer Protocol”, RFC 2821, April 2001

[11] T. Berners-Lee, R. Fielding, and H. Frystyk. “Hypertext Transfer Pretecel-HTTP/1.O”, RFC 1945, May 1996.

[12] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, “RTP: A transport protocol for real-time applications”, RFC 1889, Jan 1996.

[13] M. Handley and V. Jacobson, “SDP: Session Description Protocol”, RFC 2327, April 1998.

[14] Miikka Poikselkã, George Mayer, Hisham Khartabil, Aki Niemi, “The IMS IP Multimedia Concepts and Services in the Mobile Domain”, June 8, 2004.

[15] P. Srisuresh, K. Egevang, “Traditional IP Network Address Translator (Traditional NAT)”, RFC 3022, 2001

[16] V. Jacobson, C. Leres and S. McCanne, “tcpdump”, LBNL, University of California, June 1997.

[17] C. Sanders, “Practical Packet Analysis: Using Wireshark to Solve Real-World Network Problems”, May 2007.

[18] A. Moore, D. Zuev and M. L. Crogan, “Discriminators for use in flow-based classification”, RR-05.13, Department of Computer Science, University of London, 2005.

[19] G. Dunteman, “Principal Component Analysis”, SAGE Publications, 1989.

[20] L. Kaufman and P. Rousseeuw, “Finding Groups in Data: An Introduction to Cluster Analysis”, New York, J.Wiley & Sons, 1990.

A Survey of Energy-Efficient and QoS-Aware Routing Protocols for Wireless Sensor Networks

*G M Shafiullah, Amoakoh Gyasi-Agyei, Peter J Wolfs
Central Queensland University, Rockhampton, QLD-4702, Australia
Email: g.shafiullah@cqu.edu.au

Abstract- *Recent developments in wireless communications have enabled the development of low-cost, low-power wireless sensor networks (WSN) with wide applicability, including environment and vehicle-health monitoring. Minimizing energy consumption and hence maximizing the life time of the network are key requirements in the design of optimum sensor networking protocols and algorithms. Several routing protocols with different objectives have already been proposed for energy-efficient WSN applications. This paper surveys a sample of existing energy-efficient cluster-based and QoS-aware routing protocols and highlights their key features, including strengths and weaknesses.*

Index Terms— Energy-efficiency, routing protocol, wireless sensor network, Clustering and QoS-aware routing.

I. INTRODUCTION

ADVANCES in embedded microprocessors, low-power analog and digital electronics and radio communications have enabled the development of small and cheap sensor nodes that make the wireless sensor technology one of the promising technologies of today. Each sensor node in a WSN has wireless communication capability and some level of intelligence to collect data and route them to the sink node to enhance operational efficiency. A sensor node in a WSN comprises mainly *a sensing unit*, *a processing unit*, *a radio transceiver unit* and *a power unit*. Due to sensor nodes' placement with respect to the operational environment and quantity, it is not usually easy or even feasible to replace individual sensor nodes or their energy frequently. Sensor network applications require long lifetime, data accuracy, and energy efficiency. Hence the network needs to ensure quality of service (QoS) besides ease of deployment, energy efficiency and low cost. One of the major design goals of WSNs is reliable data communication under minimum energy depletion to extend the lifetime of the network. This may be achieved via aggressive energy management techniques. Owing to their poor energy conservation, traditional routing protocols are not suitable for WSN applications. It is highly desirable to employ an energy-efficient route discovery and data relaying techniques to transfer data between the sensor nodes and the base station (BS), the data aggregation point. Some of the routing challenges and design issues that affect the routing process in WSN are: node deployment, data reporting method, node/link heterogeneity, fault tolerance, scalability, transmission media, data aggregation, connectivity, coverage and QoS [1],[2],[3].

In this paper, we survey energy-efficient routing protocols for WSNs, some of which exploits the concept of clustering. WSNs have a wide applicability including railway infrastructure monitoring [4]. The rest of this paper is organized as follows. Section II classifies routing protocols. Section III discusses existing popular cluster-based routing protocols for WSNs. Section IV discusses recent QoS-aware routing protocols for WSNs. Section V concludes the article.

II. ROUTING PROTOCOLS FOR WIRELESS SENSOR NETWORKS

Data gathering and processing are important applications of sensor networks and all data from individual sensor nodes need to be sent to the BS, where the end user can access the data. Several routing techniques introduced for WSNs' are based on special features like, data aggregation and processing, in network processing, clustering, nodes' role assignment, and position of node. Therefore, routing protocols for WSNs can be classified into data-centric or flat-based, hierarchical or cluster-based and location-based, depending on the network structure. Routing protocols can also be divided into multipath based, QoS-based, query-based, and coherent-based, depending on how the protocol operates. This is illustrated in Fig.1.

All nodes in a flat routing protocol are assigned equal roles or functionality and the nodes collaborate to perform the sensing tasks. The BS sends queries to certain regions within the WSN and awaits data from the sensors located in that region. SPIN [5] and directed diffusion [6] are examples of flat routing protocols. Location or position information of sensor nodes is essential to calculate the distance between neighboring nodes. If the location of sensor nodes are known then data transmission only occurs on that region to reduce the number of transmissions. GAF [7] and GEAR [8] are examples of location-based energy-efficient routing protocols. In the cluster-based routing approach the network is divided into clusters and nodes play multiple roles in the network. Nodes in a cluster communicate with a cluster-head (CH) and the CH transmits the data to the global BS. This reduces the transmission range of normal nodes to conserve energy.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 352–357.

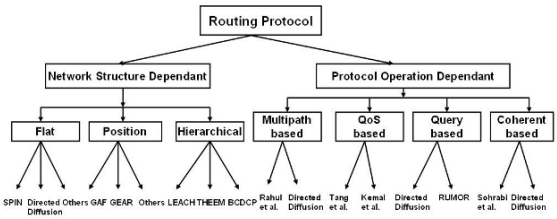


Fig. 1: Routing Protocols for Wireless Sensor Networks

The routing protocol that uses multipath instead of a single path to increase reliability is called multipath routing protocol. Directed diffusion [6] and proposal [9] are examples of multipath routing. QoS-aware routing protocols minimize the end-to-end delay, energy consumption and bandwidth requirements of WSNs during data communication. In query-based routing, the destination nodes initiate a query for data from a node through the network. The nodes having the answer to the query send the data back to the parent nodes that initiated the query. Examples of query-based routing protocols are directed diffusion [6] and RUMOR [10]. Based on data processing techniques, routing protocols for WSNs comprise coherent based and non-coherent-based. Proposal [11] and directed diffusion [6] are examples of coherent and non-coherent routing protocols [2], [3]. In this paper we highlight cluster and QoS-aware routing protocols.

III. CLUSTER-BASED ROUTING PROTOCOLS

A single-gateway architecture is not suitable for a large sensor network covering a huge area of interest. Clustering approach is a suitable solution for a sensor network with heavy load and wide coverage area. Cluster-based routing reduces energy consumption within a cluster, performs data aggregation and fusion to reduce the number of transmitted messages to the BS. Cluster-based routing, due to its salient features, is a promising area and hence has been widely studied on several fronts including: cluster formation, CH election, node to CH data communication and CH to BS data communication [2], [3]. Below we review some popular cluster-based routing protocols.

A. PEGASIS

Power-efficient gathering in sensor information systems (PEGASIS) [12] is a near-optimal chain-based protocol in which sensor nodes are organized to form a chain and each node transmits/receives data to/from its closest neighbour. Gathered data move from node to node, get fused and finally a designated node transmits the data to the BS. The nodes in the network are organized to form a chain, which can either be computed in a centralized manner by the BS or broadcast to all nodes. It could also be accomplished by the sensor nodes themselves by using the greedy algorithm. The chain construction process starts with the farthest node from the BS. The farthest node is chosen to ensure that nodes farther from the BS have close neighbours. As in the greedy algorithm, the neighbour distances increase gradually since nodes already in the chain cannot be revisited. The chain is reconstructed in the same manner to bypass any dead node in the network.

In each round, each node receives data from its neighbour node, fuses the data with its own and then transmits to the neighbour node in the chain. In turn, nodes transmit fused data to the BS. In this approach in round i the total number of nodes used is $i \bmod N$ (N represents the number of nodes) that transmit to the BS. Therefore, the leader in each communication round is at a random position in the chain and nodes die at random locations, making the sensor network robust to failures. Each round of data collection can be initiated by the BS with a beacon signal which synchronizes all sensor nodes. A time slot approach for transmitting data is employed as all the nodes in the chain know their positions. Data fusion takes place at every node except the end nodes in the chain. Therefore, in PEGASIS each node, except the two end nodes and the leader node, receive and transmit one data packet in each round and becomes the leader once every N rounds. There are some nodes which are far away along the chain and dissipate more energy in each round. These nodes cannot act as leaders in this protocol. Reference [12] introduce a chain-based binary scheme for sensor networks with CDMA nodes and a chain-based 3-level scheme for sensor networks with non-CDMA nodes to reduce the energy and delay for data gathering in WSNs.

B. LEACH

Low-energy adaptive clustering hierarchy (LEACH) [13] combines cluster-based routing and MAC protocol to serve wireless micro-sensor networks which achieve low energy dissipation and latency without sacrificing application-specific quality. The key design goals of LEACH are: adaptive cluster formation, energy-efficient MAC, and application specific data processing. LEACH attempts to distribute evenly the energy load among the network nodes by randomly rotating the cluster-head among the nodes. It uses localized coordination to enable scalability and robustness for dynamic networks. Nodes in a WSN organize themselves into local clusters, with one node acting as the CH. The CH is responsible for receiving the data from all other nodes within the clusters and aggregates the received data into a smaller set of information that describes the events the nodes are sensing. LEACH operations are divided into rounds and each round has two phases: set-up and steady-state. Clusters are organized in the setup phase, while data communication takes place in the steady state phase. A distribution algorithm is used to form the cluster and the CH where nodes make autonomous decisions without any centralized control. Each non-CH node determines to which cluster it belongs by choosing the CH that requires the

minimum transmission energy, based on the received signal strengths of the advertisements from the CHs. The CH sets up a TDMA schedule for all nodes to ensure collision-free transmissions via scheduling. At the beginning of the steady state phase nodes send data to the CH during their assigned time slot. After receiving all data from the nodes, the CH starts to aggregate data. Finally, the resultant data are sent to the BS using a fixed spreading code and CSMA protocol.

Reference [13] also introduces a centralized control algorithm called LEACH-C to form clusters by dispersing up the CH nodes in the network. Each node sends a message including its ID, location and energy level to the BS. The BS computes the energy levels of nodes and the node which has above average energy level can act as the CH. Simulated annealing algorithm [14] is used to find the CH for the rounds and the associated clusters. Minimum energy is required to transmit information from node to the CH. After formation of the cluster and the CH, the BS transmits this information to all nodes by sending a broadcast message that contains CH ID to each node. If the CH ID matches with a node's ID, then it acts as the CH, unless the node determines its TDMA slot for data transmission. Steady state phase of LEACH-C is the same as LEACH. LEACH-F is another protocol proposed in [13], in which the cluster is fixed throughout the network but the CH selection is rotational.

C. BMA

Bit-map assisted (BMA) [15] is a schedule-based MAC protocol for intra-cluster communications to reduce the energy wastes due to idle listening and collisions. In BMA sensor nodes forward data to the cluster head (CH) only if a significant event has been observed, i.e., dealing only with event-driven networks. BMA protocol reduces energy consumption occurring in conventional TDMA systems due to idle listening in the absence of data in any node in their allocated scheduled time slots. The cluster setup phase is identical to the LEACH [13] protocol. Each node in the cluster keeps its radio on in the contention period and transmits a 1-bit control message during its allocated slot if it has data to transmit, otherwise the slot remains empty. The nodes which have data to transmit are called source nodes and the rest are non-source nodes. After completion of the contention period the CH sets up and broadcasts a transmission schedule only for the source nodes. During data transmission period, each source node turns on its radio and sends its data to the CH over its allocated slot and turns off its radio at all other times. Non-source nodes keep their radios off during the data transmission period. If there is no source node, the system proceeds directly to an idle period. Analytical models are developed for BMA, TDMA and energy-efficient TDMA (E-TDMA) to validate their energy-efficient features. BMA is an energy-efficient intra-cluster low-latency MAC protocol with low complexity and reduced energy wastes in idle listening.

TABLE: 1
COMPARATIVE ANALYSIS OF CLUSTER-BASED ROUTING PROTOCOLS

Proposal	Strengths	Weaknesses
PEGASIS [12]	Distributing the energy load among the nodes increases the lifetime and quality of the network. Low energy consumption transmission. PEGASIS outperforms LEACH in energy consumption.	High latency for distant nodes on the chain and overhead associated due to topology adjustment. If there are *N* nodes in the network the BS makes *N* data jumps to acquire all the data from the network which introduces time delay.
LEACH [13]	Overall throughput of the network increases as latency is reduced and system life increases. Collision and interference are overcome by using TDMA and CDMA protocol respectively. Data aggregation mechanism reduces the amount of data that must be transmitted to a BS, hence reduces energy consumption. LEACH outperforms direct approach and MTE routing protocol.	Due to its distributed cluster formation algorithm, it cannot ensure the coverage of entire network. TDMA schedule introduces time delay. Comparatively it has more setup overhead.
BMA[15]	BMA has low complexity, reduces energy wastes due to idle listening and low packet latency. BMA outperforms TDMA and E-TDMA in low and medium traffic loads.	BMA is suitable only for low traffic, i.e., relatively few sensor nodes per cluster. Data may arrive in empty node at any time during node to the CH data transmission.
BCDCP [16]	Balanced clustering techniques, CH-to-CH routing scheme makes BCDCP an energy-efficient protocol. BCDCP reduces energy expenditure and increases network lifetime over its rivals (LEACH, LEACH-C and PEGASIS).	TDMA protocol introduces latency. BS needs more energy to perform most of the tasks.
THEEM [18]	Eliminate cluster setup overhead in each round as cluster is fixed throughout the network. TDMA schedule and CSMA techniques are used to reduce collision and interference respectively. THEEM outperforms MTE and PEGASIS routing protocol.	Fixed cluster protocol is unable to adapt to new conditions. A node may need to transmit data in a far way and needs more energy to communicate with its MH when there is another cluster's MH nearby.

D. BCDCP

Base-station controlled dynamic clustering protocol (BCDCP) [16] is a centralized clustering-based routing protocol in which network lifetime is improved by

distributing the energy dissipation among nodes in the network. Here the BS plays the key role to set up clusters, CH selection, routing path and other energy-intensive tasks. Key features of this protocol are: equal number of nodes in each cluster avoids CH overhead; position of CH is uniform throughout the network; and communication occurs in the manner normal node-to-CH, CH-to-CH and CH-to-BS.

Sensor nodes are grouped into clusters and work in either sensing mode or the CH mode. In sensing mode, the node transmits sensed phenomenon to the CH. There are setup and communications phases in BCDCP. The main activities of the setup phase are: cluster setup, CH selection, formation of CH-to-CH routing path and data transmission schedule. In each setup phase the BS receives information about nodes' energy levels and computes average energy level of all nodes in the network. The BS then select a set of nodes that has above average energy level and a cluster splitting algorithm is used to select CHs from this set. Balanced clustering technique is used to distribute evenly the load on all cluster heads. After selecting clusters and their CHs using the spanning tree algorithm [17], the BS forwards the lowest-energy routing path information to the sensor nodes with information on the cluster and the CH. TDMA scheduling approach is used to transmit data from sensor nodes to the CH. Data gathering, fusion and data routing activities are involved in the data communication phase. To prevent radio interference caused by neighbouring nodes, BCDCP uses CDMA codes during CH-to-CH multihop routing [16].

E. THEEM

Two-hop energy-efficient mesh protocol (THEEM) is a hierarchical cluster-based energy-efficient protocol. The main goals of this protocol are to achieve energy-efficiency, high QoS, low latency, and equilibrium energy dissipation. The network is divided uniformly into clusters or meshes by using a centrally organized method to eliminate the mesh setup overheads. Residual energy of each mesh is calculated at the beginning of each round and, based on the energy levels the mesh-head (MH) is chosen. The MH is rotated and any node can act as MH.

In each round there are setup and assignment phases in THEEM. Nodes send out short message with their location and energy level to the BS at the initial stage of the setup phase. The BS broadcasts control packets to the network and all sensor nodes synchronize their clocks and update their tables. Assignment phase is composed of the start beacon, energy reporting form and control packet. The start beacon is sent by the BS to inform the network about the start of the current round. After receiving node location and energy level the BS selects the MH for the current round according to their residual energy. The BS then makes a TDMA schedule for each node and broadcasts a control packet to the network. THEEM introduces a two-hop scheme to transmit data from a node to the MH and hence reduces energy consumption. Meshes are divided into two portions, the portion which is far from the MH is called source node and the portion which is closer to the MH is called intermediate node. As MH is changed with rounds, source and intermediate nodes also change with rounds. In the first hop, source nodes send data to intermediate nodes using the minimum transmission energy (MTE) [19] routing approach. In the second hop, the intermediate node fuses the data and sends it to the MH. A multi-hop scheme is used to transmit data from the MH to the BS. CDMA code is used to reduce interference between meshes. The farthest mesh layer transmits data to the adjacent layer and this layer sends accumulated data to the next adjacent layer and finally to the BS. Reduced overhead and CH-to-CH routing makes this an energy-efficient protocol [18].

IV. QoS-AWARE ROUTING PROTOCOLS

The key design goals of QoS-aware routing protocols are to ensure optimized QoS metrics such as delay bound, energy efficiency, low bandwidth consumption while achieving energy efficiency in WSNs applications. Few recent QoS-aware protocols are surveyed in the following sections.

A. QoS-Aware Routing

Reference [20] has introduced a cluster-based linear WSN topology to support QoS and optimal energy allocation in both homogenous and heterogeneous WSNs. A multihop WSN is considered in which each CH acts as the coordinator of a cluster. Each cluster is managed by a single server fixed rate (SSFR) node with finite capacity and different data arrival rates. The CH communicates with its adjacent cluster's CH and hence data is transmitted from one cluster to another until it reaches the sink. In this work the CH is modelled using M/M/1/K queuing system in a first come first served (FCFS) manner. The CH monitors the data loss probability (DLP) and controls data generation rate to ensure QoS requirements. DLP increases with increasing traffic load. Therefore, by continuously adjusting the data generating rate at individual clusters the DLP is optimized. Optimal energy allocation algorithm makes all clusters have the same exhaustion time and hence reduces the problem of early die in the cluster closest to the sink. The clusters with high-density nodes decrease data generating rate per node to save energy. Transmission energy is optimized by adjusting the radio range. The clusters near the sink relay data even shorter distances to reduce their energy dissipation while clusters at the rear relay over longer distances to improve the relaying efficiency. Numerical analysis shows that the DLP increases with increasing of data generating rate and drastically increases at the clusters closer to the sink. The source-to-sink delay (SSD) also increases with data generating rate but the increasing rate gradually slows as the data generating rate increases.

B. Energy and QoS-Aware Routing

Reference [21] presents a novel energy-aware protocol for routing of delay-constrained data generated by imaging

sensors. Collecting sensed imaging data requires both energy and QoS-aware routing to achieve energy efficiency of the sensor networks. Main aim of this protocol is to find a delay constrained path with minimum cost. Communication cost, nodes' energy stock, error rate, and other communication parameters determine the link cost. M/M/1 queuing model is used for analysis and separate queues are used for real-time and non-real-time data. Classifier checks the incoming packet and sends to the appropriate queue. Scheduler determines the order of packets to be transmitted from the queues. End-to-end delay of a path is measured by using queuing model. The bandwidth ratio (r) represents the amount of bandwidth to be allocated both to the real-time and non-real-time traffic on a particular link. The queuing delay depends on this bandwidth ratio r. Optimal selection of r can meet end-to-end QoS requirements. After calculating costs of each link an extended version of Dijkstra algorithm [22] is used to find the least-cost path which meets the end-to-end delay requirements. Single-r and multi-r mechanisms are introduced to calculate end-to-end delay for a particular path and to find out an optimal value for r. The effectiveness of the protocol is validated through simulation and metrics considered to evaluate performance are: time to first node to die, average lifetime of a node, and average delay per packet and network throughput.

C. ESRT

Event-to-sink reliable transport (ESRT) is a transport protocol for WSN that seeks to achieve reliable event detection with minimum energy expenditure and congestion resolution. The reliability of event feature is measured by the number of received data packets which is closely related to the amount of information acquired by the sink for the detection and extraction of event features. The event is reliable if the observed event reliability is greater than the desired event reliability. The reporting frequency rate f of a sensor node is the number of packets sent out per unit time by that node. The transport problem in WSN is to configure the reporting frequency f, of source nodes so as to achieve the required event detection reliability R at the sink with minimum resource utilization. The data generated by the sensors are temporally correlated to tolerate individual packet losses to the extent where the distortion D_i observed when the event features are estimated at the sink does not exceed a certain distortion bound D_{max}. Reporting frequency f can control the amount of traffic injected to the sensor field while regulating the number of temporally-correlated samples taken from the phenomenon. The desired event distortion is achieved by selecting a small f.

TABLE: II
COMPARATIVE ANALYSIS OF QOS-AWARE ROUTING PROTOCOLS

Proposal	Strengths	Weaknesses
Tang et al. [20]	Optimal energy allocation scheme increases network lifetime and ensures QoS under heavy traffic conditions. This protocol saves energy of the network and supports QoS.	FCFS queuing essentially relegates all congestion control to the sources and completely depends on traffic arrival time. Inappropriate control of data generating rate increases energy consumption.
Akkaya et al. [21]	Enhanced throughput and average delays. By adjusting service rate the throughput of non-real-time data is increased. Good QoS and energy efficiency.	Inappropriate adjustment of packet service rate introduces delay per packet and wastes energy. Energy consumption is comparatively high in multi-r mechanism.
ESRT [23]	Reliability and congestion control are achieved by changing the reporting rate of nodes. Power consumption reduces with time and minimizes overall energy expenditure by maintaining the operation in the OOR state.	Regulating all sensors to have the same reporting rate may not work well with heterogeneous sensors. Congestion in WSN is not only caused by frequent sensor reporting.

Relationship between the observed reliability r and the reporting frequency f shows that the reliability increases linearly with source reporting rate f, until a certain $f = f_{max}$, beyond which the reliability drops as packets are dropped due to congestion. For $f > f_{max}$ the behavior is not smooth and drop in reliability due to network congestion is more significant with increasing n. Normalized reliability = r/R = 1 gives a suitable point known as optimal operating region (OOR) to achieve protocol's goal. Five characteristic regions are identified: (NC, LR), (NC, HR), (C, HR), (C, LR), (OOR) considering congestion and reliability. Details of these regions are given in [23].

The primary motive of ESRT is to achieve and maintain operation in the state OOR. So the aim is to configure the reporting frequency f to achieve the desired event detection accuracy with minimum energy expenditure. Reference [23] uses a congestion control mechanism that serves the dual purpose of reliable detection and energy conservation. An efficient transport protocol should keep track of the reliability observed at the sink and accordingly configure the operating point. They developed methodology both for single event and multiple events. The sink must accurately detect whether the flows generated by these multiple events pass through any common nodes functioning as router [23].

V. CONCLUSION

Owing to their versatility and efficient features, routing protocols play an important role in the operation of an energy-efficient WSNs. Therefore, they are promising research area. Based on application area and network architecture, several routing protocols have been introduced that ensure energy-efficiency to prolong the network life. This paper presents a concise survey of cluster-based and QoS-aware routing protocols. Cluster-based routing protocols divide the sensor networks into several clusters. Nodes serving as CHs have higher energy than the normal nodes as they perform data aggregation and data processing. Cluster formation, CH election and data communications are designed to keep the power consumption and latency as minimum as possible. QoS-aware routing protocols are designed to minimize bandwidth consumption and latency. Minimum-cost paths are selected for data communications to ensure integrity.

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," IEEE Comm. Magazine, pp. 102–114, Aug'02.

[2] J. N. Al-Karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: A survey," IEEE Wireless Communications, Dec'04.

[3] K. Akkaya and M. Younis, "A survey on routing protocols for wireless sensor networks," Ad Hoc Networks, vol. 3, Nov'03.

[4] G. M. Shafiullah, A. Gyasi-Agyei, and P. Wolfs, "Survey of wireless communications applications in the railway industry," in Conf. on Wireless Broadband and Ultra Wideband Comm., Sydney, Aug'07.

[5] W. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks," in 5th ACM/IEEE mobicom, Seattle, WA, Aug'99, pp. 174–185.

[6] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion: a scalable and robust communication paradigm for sensor networks," in 6th ACM/IEEE Mobicom'00, Boston, MA, Aug'00, pp. 56–67.

[7] Y. Xu, J. Heidemann, and D. Estrin, "Geography-informed energy conservation for ad hoc routing," in 7th ACM/IEEE Mobicom'01, Rome, Italy, Jul'01, pp. 70–84.

[8] Y. Yu, D. Estrin, and R. Govindan, "Geographical and energy-aware routing: a recursive data dissemination protocol for wireless sensor networks," UCLA Computer Science, Tech. Rep. TR-01-0023, May'01.

[9] C. Rahul and J. Rabaey, "Energy aware routing for low energy adhoc sensor networks," vol. 1, Orlando, Florida, USA, Mar'02, pp. 350–355.

[10] D. Braginsky and D. Estrin, "Rumor routing algorithm for sensor networks," in Int. conf. on Distributed Computing Systems, Nov'01.

[11] K. Sohrabi and J. Pottie, "Protocols for self-organaization of a wireless sensor network," IEEE Personal Comm., vol. 7, pp. 16–27, 2000.

[12] S. Lindsey, C. Raghavendra, and K. M. Sivalingam, "Data gathering algorithms in sensor networks using energy metrics," IEEE Transaction on Parallel and distributed systems, vol. 13, pp. 924–935, Sep'02.

[13] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," IEEE Trans. on Wireless Comm., vol. 1, pp. 660–669, Oct'02.

[14] T. Murata and H. Ishibuchi, "Performance evaluation of genetic algorithms for flowshop scheduling problems," in Proc. of 1st IEEE Conf.on Evolutionary Computation, vol. 2, Jun'94, pp. 812–817.

[15] G. Y. Lazarou, I. Li, and J. Picone, "A cluster-based power-efficient MAC scheme for event-driven sensing applications," Ad Hoc Networks, vol. 1, May'06.

[16] S. D. Muraganathan, D. C. MA, R. I. Bhasin, and A. O. Fapojumo, "A centralized energy-efficient routing protocol for wireless sensor networks," IEEE Radio Communications, vol. 43, pp. S8–S13, Mar'05.

[17] H. Shen, "Finding the k most vital edges with respect to minimum spanning tree," in Proc. IEEE Nat'l Aerospace and Elec. Conf., vol. 1, Jul'97, pp. 255–262.

[18] Y. Liu, "A two-hop energy-efficient mesh protocol for wireless sensor networks," Ph.D. dissertation, Department of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, Canada, 2004.

[19] M. Ettus, "System capacity, latency, and power consumption in multihop-routed SS-CDMA wireless networks," in Radio and wireless conf.(RAWCON), Colorado Springs, CO, Aug'98, pp. 55–58.

[20] S. Tang and W. Li, "QoS supporting and optimal energy allocation for a cluster based wireless sensor network," Computer Communications, vol. 29, pp. 2569–2577, Mar'06.

[21] K. Akkaya and M. Younis, "Energy and qos aware routing in wireless sensor networks," Cluster computing, vol. 8, pp. 179–188, 2005.

[22] E. de Queirs Vieira Martins, M. M. B. Pascoal, and J. L. E. dos Santos, "The k shortest paths problem," CISUC, Tech. Rep., Jun'98.

[23] O. B. Akan and I. F. Akyildiz, "Event-to-sink reliable transport in wireless sensor networks," IEEE/ACM transactions on networking, vol. 13, pp. 1003–1016, Oct'05.

Stepping-Stone Intrusion Detection Using Neural Networks Approach

Han-Ching Wu and Shou-Hsuan Stephen Huang
Department of Computer Science, University of Houston,
Houston, TX 77204, USA
Email: {hcwu, shuang}@cs.uh.edu

Abstract - **Attackers on the Internet typically launch network intrusions indirectly by creating a long connection via intermediary hosts, called stepping-stones. One way to detect such intrusion is to check the number of intermediary hosts. Neural networks provide the potential to identify and classify network activity. In this paper, we propose an approach to stepping-stone intrusion detection that utilizes the analytical strengths of neural networks. An improved scheme was developed for neural network investigation. This method clustered a sequence of consecutive Round-Trip Times (RTTs). It was found that neural networks were able to predict the number of stepping-stones for incoming packets by the proposed method without monitoring a connection chain all the time.**

Keywords - Stepping-stone detection, neural networks, round-trip time, packet matching.

I. INTRODUCTION

Due to the prevalence of web services and electronic commerce, the issue of network security and intrusion detection has become more and more important. Network intruders often launch attacks through a long connection chain via intermediary hosts, called *stepping-stones* that they previously compromised rather than directly attack their targets [1]. This widely used technique by intruders is called *stepping-stones intrusion*. One way to detect stepping-stone intrusion is to estimate the downstream length of the connection chain. It has been observed that the round-trip time (RTT) of a packet in the connection chain is a function of the length of the chain [2].

Several approaches have been developed to detect stepping-stone intrusion in recent years. One earlier method with a content-based algorithm was proposed by Staniford-Chen and Heberlein [3]. Another line of work studied correlation of streams based on connection timings. Zhang and Paxson [4] presented a method to investigate correlation of connections and to match the on- and off-periods. A deviation-based algorithm was reported by Yoda and Etoh [5]. Donoho et al. [6] used wavelet transforms to examine the maximum tolerable delay. Wang and Reeves [7] suggested a watermark-based approach to study the perturbation for packet inter-arrivals. He and Tong [8] gave two activity-based algorithms with bounded memory or bounded delay perturbation. Blum et al. [9] proposed algorithms to study the monitoring time and false positive rate with and without chaff perturbation based on computational and random walks theories. Wu and Huang [10] proposed an algorithm to identify stepping-stones under different chaff rates.

When compared to the studies mentioned above, neural networks may replace conventional algorithms using statistical methods and are powerful techniques for the simulation of intrusion detection. Cannady [11] investigated a network-based neural network detection system in which packet-level network data were retrieved and then classified based on nine packet characteristics. Ghosh et al. [12] built a neural network to analyze program behavior profiles instead of user behavior profiles. Some other groups (Lippman and Cunningham [13] and Ryan et al. [14]) created keyword count based misuse detection systems with neural networks.

Yang et al. [15] proposed a clustering method to identify the length of a connection chain. However, the method in [15] was restricted to the assumption that one is able to monitor a connection chain during the whole connection session. Wu and Huang [16] also suggested a method to solve this problem using packet variables of an *individual* packet to predict the number of stepping-stones.

In this paper, we propose an improved approach from [16] to overcome this problem by developing *Scheme G* with neural networks. After the neural networks were trained, new coming packets could be treated as testing data, so that we did not need to monitor a connection chain all the time and the number of stepping-stones could be obtained. To take into account RTTs as the only variable, we clustered a sequence of consecutive RTTs as a *group* to estimate the downstream lengths. The results were compared and discussed with [16]. It was noted that references [11] to [14] did not study the effect of transfer function and learning rule on the predicted results, and this effect is studied in this paper.

The rest of the paper is organized as follows. The stepping-stone intrusion detection is outlined in Section II. Section III summarizes the principle of neural networks. Section IV develops a scheme for neural networks investigation. The experiments and the predicted results are analyzed in Section V. Finally, the conclusion is made in Sections VI.

II. STEPPING-STONE INTRUSION DETECTION

A. Distinguishing Intruders from Normal Users

Network intruders often launch attacks by creating a long connection chain via intermediary hosts (see Figure 1), called *stepping-stones*, in order to evade detection [1]. A legitimate user or application normally accesses the target host through a

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 358–363.

direct connection and seldom through a long chain with more than two connections. Therefore, an effective way to detect the stepping-stone intrusion is to estimate the downstream length from the monitoring computer to the victim computer. The longer a downstream chain length is, the more likely an intrusion exists [2].

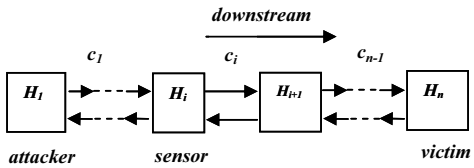


Fig. 1. A Connection Chain

A connection c_i is a connection from computer host H_i to host H_{i+1}. Figure 1 shows a typical connection chain from c_1 to c_{n-1}. An attacker in computer H_1 launches an attack to the victim computer H_n via the stepping-stones H_2, …, H_i, …, H_{n-1}. The host H_i is called *sensor* where we put the detecting program to monitor the connection chain. The direction toward the victim host from the sensor is called downstream, and the downstream length denotes the number of connections.

B. Packet Round-Trip Times (RTTs)

We use Send to indicate a TCP packet downstream and Echo as the TCP packet upstream. If one monitors a downstream connection chain at a sensor, one is able to collect all the Send and Echo packets of the connection. Then, we compute the time interval between each send and its matched echo packet; this time interval is the Round-Trip Times (RTTs) of the packet.

A packet RTT may be influenced by various factors and is mainly decided by four kinds of delays on each connection traversed by the packet. For a single connection, a RTT can be expressed as

$$RTT = 2(d_1 + d_2 + d_3 + d_4) \quad (1)$$

Where the four delays include propagation delay (d_1), processing delay (d_2), queuing delay (d_3) and transmission delay (d_4) [17].

Since the time d_1 is determined by physical distance. If we do not consider the bandwidth problem, d_1 is relatively stable for different packets. Let R_0 be the invariant delay, (i.e. $2d_1$) and ΔR be variant delays $2(d_2,+d_3 + d_4)$, then we have

$$RTT = R_0 + \Delta R \quad (2)$$

If there are k connections (or k hosts) in one connection chain, then let ΔT represent the varying part of all queue delays along the connection chain [17]:

$$\Delta T = \sum_{i=1}^{k} \Delta R_i \quad (3)$$

C. TCP packet matching

To compute the RTTs, we need to consider the problem of TCP packet matching. The matching algorithm may become complicated because the Send and Echo packets may be a many-to-many relationship, not just one-to-one. If there were only one-to-one correspondence, for each Echo packet, we could match it with the first unmatched Send packet; then, this is simple matching. In fact, there exist many simple matching packets in our experiments. However, there are still some one-to-many and many-to-one cases in our experiments. Therefore, we need to use a more sophistic matching algorithm to calculate the RTTs.

```
Initialize Send queue;
while(there are more packets){
   read next packet P;
   if P is a send packet {
      calculate TimeGap;
      if TimeGap > threshold
         reset Send queue;
      else
         insert P to Send queue; }
   else if P is an echo packet {
      if (P.size is large) do nothing;
      else {
         S = dequeue(Send);
         if((S.ack = P.seq)and(S.seq < P.ack){
               Match P and S;
               Compute RTT between P and S; }}}
}
```

Fig. 2. Matching Algorithm

Figure 2 shows the matching algorithm we used that is modified from the matching algorithm in [18]. We read each new unmatched Send packet and put it into a Send queue. While reading a new Echo packet, there are two types of echo packets. The first type carries bounced character (thus has a very small size, called *minimum size*); the second type carries response from the host with size larger than that of the first type. We first check if echo packet is larger than the minimum size. If yes, we filter it and read next one. If not, we continue the matching process: (1) if there is only one Send packet in the Send queue, match both; (2) if more than one unmatched Send packets are in the Send queue, we match Echo with the first Send. The conditions Send.Ack = Echo.Seq and Send.Seq < Echo.Ack are checked to ensure the correct matching.

D. Number of Stepping-Stones

Suppose there is only one connection in a downstream chain, each of the computed RTTs should be very close. We cannot expect all RTTs of the same connection to be the same since there are several reasons to make RTTs vary, such as the network traffic fluctuation [19]. However, most RTTs vary only slightly if they are collected from the same connection chain in a small time interval. Those RTTs have a clear lower bound. Each time one more connection is extended, we get one more RTT level by matching Send and Echo packets and compute RTTs. Hence, estimating the number of stepping-stones (i.e. downstream length) simplifies to counting the RTT levels, which can be examined by neural networks.

III. NEURAL NETWORKS

Neural Networks (NN) or Artificial Neural Networks (ANN) are usually considered as an effective approach for classification [20]. The advantages of NN in the detection of instances are flexibility and the inherent speed. NN also possesses the ability to analyze the non-linear data sets with

multi-variables [12]. As mentioned in Section 1, [15] was restricted to the assumption that monitoring a connection chain all the time is required. Using neural networks we propose a novel approach to overcome this problem by developing a new scheme. After the neural networks are trained, new coming packets can be treated as testing data, so that we do not need to monitor a connection chain all the time and the number of stepping-stones can be obtained. Once the neural networks have been generated, they have to be re-generated for different data sets. Fortunately, we can utilize the advantages of fast regeneration and analytical strengths of neural networks for new connection chains.

In this study, we use a back-propagation neural network [20], in which a *transfer function* is typically a non-linear function. The weighted inputs are summed, and passed through a transfer function, which scales the output to a fixed range of values and broadcasts to all of the neurons in the next layer. When designing a neural network, the initial transfer function chosen applies to each layer of the neural network. In this paper, three various transfer functions are used. They are Sigmoid, TanH and Sine functions. In addition, three *learning rules* for the weight updates are also considered. They are Delta, Norm-Cum-Delta (NCD) and Delta-Bar-Delta (DBD) rules. The effect of transfer function and learning rule on the predicted results is studied in Section 5.

IV. SCHEME DEVELOPMENT

We developed a new scheme with neural networks to improve the method proposed in [16], which we called the *Individual Packet* Scheme or *Scheme I*. The Scheme *I* basically was an intuitive method to study the effects of related variables in the TCP packet data and each Individual network packet is examined. This method used packet variables of an individual packet as the input layer to predict the number of stepping-stones. In this paper, the new approach (see Figure 3) is an alternative method to investigate the effect of RTT fluctuation of packets, which clustered a sequence of n consecutive RTTs as a group and generates the packet sets for the input layer to estimate the downstream lengths. The reason for doing so is to train neural networks by taking into account the fluctuation of the RTTs. This new approach is called the *RTT group* scheme or *Scheme G*.

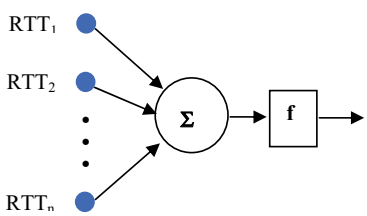


Fig. 3 RTT group scheme: with a sequence of RTTs as input

In our experiments, three preprocessing steps were conducted to prepare the data for use in the training and testing of the neural networks development. The first preprocessing step used a Packet-Capture program to collect needed TCP packets. Only Send and Echo packets were kept and Ack packets were filtered out.

The second preprocessing step ran the Matching program (see Figure 2) to compute the RTTs of each matched Send and Echo packet pair. Then every n consecutive RTTs were grouped as a row to generate the packet sets used for the input layer. For example, if there were 400 packet RTTs and let n equal 10, we divided the data into 40 rows. Then each row of RTTs was fed into the input layer of the neural networks.

The third preprocessing step was to develop the neural networks using the data set created from the second step. Then the developed neural networks could be used to identify the intrusion detection. Since we observed that RTTs were the most significant variable that affects the level of the packets, we applied this new method to check if we can estimate the downstream length simply based on the RTTs.

B. Development of Neural Networks

Neural Works Professional II/PLUS (NWorks) [21] provided a complete neural network development environment. Collected data were used for training and testing simultaneously. Using NWorks, the algorithm to build a network for stepping-stone intrusion detection is explained as follows:

(1) Network creation– *RTT group* scheme was created.
(2) Instruments setting–to send diagnostic information from a network to the screen, a file, or a user-written program.
(3) Network IO setting–to be used during learning and testing.
(4) Network initialization–to initialize the weights in all layers to random values between the low and high values specified in each layer.
(5) Network learning–the effects of both transfer functions (Sigmoid, TanH and Sine) and learning rules (Delta, NCD and DBD) were studied.
(6) Network testing–the network was trained using the trained cases, and then the test cases serve as a way of measuring network performance. Since the numbers of stepping-stones were known, the predicted values during testing were used to measure the network performance.

V. EXPERIMENTS AND RESULTS

A. Experimental Setting

We performed several tests to demonstrate the applicability of our neural networks. Due to the limitation of space, only the following selected subsets of the tests were reported. Figure 4 shows the two experimental connection chains in this paper. Group *A* included three LAN connection chains–*a1*, *a2* and *a3*; and Group *B* included three Internet connection chains–*b1*, *b2* and *b3*. The hosts, *A08*, *A09*, *Bay* and *L01*, were all located on UH campus; and *WI* and *NJ* were the off-campus hosts located in Wisconsin and New Jersey. Since all hosts on campus in Group *A* were LAN connections, they were short connections and the computed RTTs were also small. In contrast, in Group *B*, the connections were longer, so the computed RTTs were also larger. The monitoring program was installed in host *A08* to collect the packets for all experiments. In each group, the number of hosts and downstream length was from one to three.

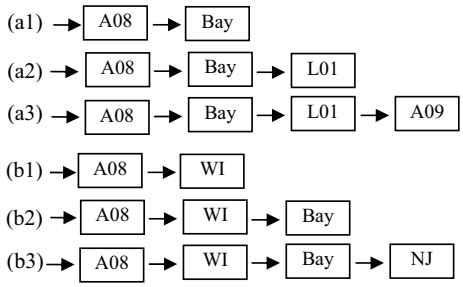


Figure 4 Two experimental connection chains: Group *A* – LAN chains (a1, a2, a3), Group *B* – Internet chains (b1, b2, b3)

B. Experimental Results

The method *Scheme G* grouped a sequence of consecutive RTTs and created the packet sets as the inputs to estimate the downstream lengths. We observe from [16] that RTT was the most significant variable to the prediction of the level of the packets. Because the local network traffic was lighter and more stable, we expected the fluctuation in Group *A* would be less than that in Group *B*. We studied the effects of both transfer functions and learning rules on the prediction using the True-and-Predicted figures, Root-Mean-Square (RMS) of errors and correct rates.

To investigate the effect of transfer functions, all the three transfer functions with the same learning rule of Delta were used. When Figure 5(a) was compared to Figure 5(b), it could be observed that the predicted results for Group *A* were better than those for Group *B* as expected.

Since the true levels of packets were integers and the predicted levels of packets were real numbers, we rounded the predicted values to the nearest integers and compared them with the true levels. For example, the true level was 2 and the predicted level was 1.8; after rounding, both had the same level 2, so we could say the predicted level was correct. We defined the *correct rate* as the ratio of the number of correct predictions to the total number of testing data. The correct rate could represent the predicted results more clearly.

The RMS values of errors and the correct rates of prediction were shown in Tables 2(a) and 2(b), respectively. From these two tables, the following is observed:

(1) Using the Sigmoid function all three learning rules gave excellent results, i.e. the RMS of errors was about 0.02 and 0.05 for Groups *A* and *B*, respectively. And the correct rates are almost 1 for both groups.

(2) The results of the RMS of errors and the correct rates of predictions from Group *A* were slightly better than those from Group *B*. The reason was that the local network traffic load was lighter and more stable than the Internet traffic, so that Group *A* had less fluctuation than Group *B*. However, the difference of both predicted results was not very large since the hosts (WI and NJ) outside the campus were not busy and still under control. In the future, more connection chains should be included in the training and testing data sets in order to make the experiments closer to the real Internet connections.

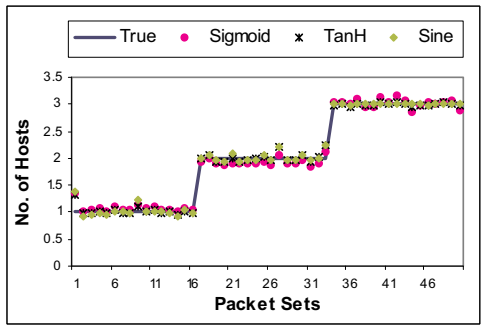


Fig. 5(a) True and Predicted results with Different Transfer Function under Delta Learning Rules for Group *A* in Scheme *G*

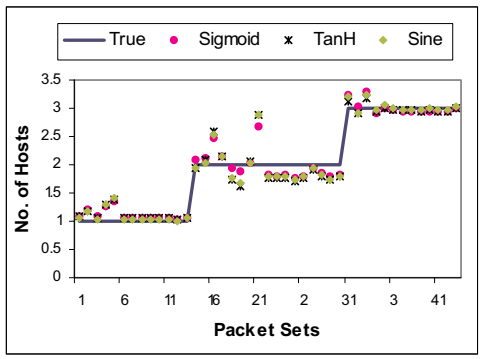


Fig. 5(b) True and Predicted results with Different Transfer Function under Delta Learning Rules for Group *B* in Scheme *G*

Table 2(a) RMS of Errors with *Scheme G*

Learn Rules	Transfer Function	Group *A*		Group *B*	
		Training	Testing	Training	Testing
Delta	TanH	0.0312	0.0602	0.1599	0.1828
Delta	Sine	0.0265	0.0666	0.1445	0.1755
Delta	Sigmoid	0.0216	0.0291	0.0495	0.0564
NCD	Sigmoid	0.0238	0.0306	0.0508	0.0515
DBD	Sigmoid	0.0275	0.0347	0.0586	0.0542

Table 2(b) Correct Rate of Prediction with *Scheme G*

Rules	Trans. Fun.	Group *A*	Group *B*
Delta	TanH	1.00	0.95
Delta	Sine	1.00	0.95
Delta	Sigmoid	1.00	0.98
NCD	Sigmoid	1.00	1.00
DBD	Sigmoid	1.00	1.00

In order to compare Scheme *G* reported in this paper with Scheme *I* reported in [16], the RMS of errors and correct rate with Scheme *I* are shown in Table 3(a) and 3(b), respectively. When Table 2(a) is compared to Table 3(a), it is found that the

RMS of errors in Scheme *I* was greater than that of Scheme *G* for Group *A*. However, for Group *B*, they were pretty close. When Table 2(b) is compared to Table 3(b), it is noted that the correct rate of prediction for Scheme *G* is slightly better than that for Scheme *I*. In short, the comparison of Schemes *G* and *I* shows that the results from Scheme *G* are better than the outcomes of Scheme *I*. The possible reason for this is that in the test data various factors such as the network bandwidth and network congestion do not significantly affect RTTs

Table 3(a) RMS of Errors with *Scheme I*

Learn Rules	Trans. Fun.	Group *A*		Group *B*	
		Training	Testing	Training	Testing
Delta	TanH	0.1136	0.1270	0.1735	0.1818
Delta	Sine	0.1098	0.1353	0.1619	0.1735
Delta	Sigmoid	0.0388	0.0538	0.0612	0.0670
NCD	Sigmoid	0.0386	0.0535	0.0592	0.0670
DBD	Sigmoid	0.1188	0.1253	0.0928	0.0983

Table 3(b) Correct Rate of Prediction with *Scheme I*

Rules	Trans. Fun.	Group *A*	Group *B*
Delta	TanH	0.99	0.96
Delta	Sine	1.00	0.96
Delta	Sigmoid	0.99	0.95
NCD	Sigmoid	0.99	0.95
DBD	Sigmoid	0.69	0.96

D. Detection Algorithm and Analysis

Even though our tests above were done as batch jobs, the algorithm could be used in real-time detection because our method predicted the number of stepping-stones by monitoring only a few packets in a connection chain. This subsection addressed issues related to real-time usage of our detection algorithm.

For Scheme *G*, the test cases were selected from a sequence of 10 packets that were all of the same chain length. In detecting intruders, we might encounter a sequence of packets selected from two different chain lengths. The accuracy for this sequence might not be as good as what is shown in Table 2. This was not a serious problem because of two reasons: (1) the probability of this cross-boundary case was relatively low, and (2) we could use a voting method to improve the accuracy of the prediction. With a window-of-three voting, we only needed to collect 12 packets to form three sequences of ten packets.

VI. CONCLUSIONS

In this paper, we studied a new scheme to examine the downstream length of the connection chain by using neural networks. The scheme clustered a sequence of consecutive RTTs. An improvement over the methods proposed recently in [15] was that we did not need to monitor a connection for the whole session. In other words, after the training of the neural networks, we only needed to collect a few packets to detect the number of stepping-stones in real-time. In this paper, we also compared the proposed Group scheme with the individual packet scheme reported in [16], and found that the performance of the Group scheme was slightly better than that of the individual packet scheme. It was also found that neural networks were capable of identifying the number of stepping-stones, and using Sigmoid transfer function and Delta learning rule generally gave very good predictions. It should be noted that once the neural networks have been generated, they have to be re-generated for different data sets. As a next step, we can utilize the advantages of fast regeneration and analytical strengths of neural networks for more complicated connection chains.

ACKNOWLEDGEMENT

Support of this research under a grant from Texas Learning and Computation Center (TLC^2) is acknowledged.

REFERENCES

[1] Yung, K.H., "Detecting Long Connecting Chains of Interactive Terminal Sessions", RAID 2002, Springer Press, Zurich, Switzerland, pp. 1-16, October 2002.

[2] Yang, J.H., and Huang, S-H.S., "A Real-Time Algorithm to Detect Long Connection Chains of Interactive Terminal Sessions", Proceedings of 3rd International Conference on Information Security (Infosecu'04), Shanghai, China, pp. 198-203, November 2004.

[3] Staniford-Chen, S. and Heberlein, L.T., "Holding Intruders Accountable on the Internet", in Proc. of the 1995 IEEE Symposium on Security and Privacy, Oakland, CA, pp. 39-49, May 1995.

[4] Zhang, Y. and Paxson, V., "Detecting Stepping Stones", in Proc. of the 9th USENIX Security Symposium, Denver, CO, pp. 171-184, August 2000.

[5] Yoda, K. and Etoh, H., "Finding a Connection Chain for Tracing Intruders", 6th European Symposium on Research in Computer Security-ESORICS 2000 LNCS-1895, Toulouse, France, pp. 31-42, October 2000.

[6] Donoho, D., Flesia, A.G., Shankar, U., Paxson, V., Coit, J. and Staniford, S., "Multiscale Stepping-Stone Detection: Detecting Pairs of Jittered Interactive Streams by Exploiting Maximum Tolerable Delay", 5th International Symposium on Recent Advances in Intrusion Detection, Lecture Notes in Computer Science 2516, New York, Springer, 2002.

[7] Wang, X. and Reeves, D., "Robust Correlation of Encrypted Attack Traffic Through Stepping Stones by Manipulation of Inter-packet Delays", in Proc. of the 2003 ACM Conference on Computer and Communications Security (CCS 2003), ACM Press, pp. 20-29, October 2003.

[8] He, T. and Tong, L., "Detecting Encrypted Stepping-stone Connections", IEEE Trans. on Signal Processing, Feb. 2006.

[9] Blum A., Song D. and Venkataraman S., "Detection of Interactive Stepping Stones: Algorithms and Confidence Bounds," in Proc. of 7th International Symposium on Recent Advances in Intrusion Detection (RAID '04), Springer LNCS 3224, pp. 258-277, 2004.

[10] Wu, H.C., Huang, S-H.S., "Detecting Stepping-Stone with Chaff Perturbation," in Proc. of AINA, International Symposium on Frontiers in Networking with Applications (FINA 2007), Ontario, CA, Vol. 1 pp. 85-90, May 2007.

[11] Cannady, J., "Artificial Neural Networks for Misuse Detection," Proceedings, National Information Systems Security Conference (NISSC'98), Arlington, VA, pp. 443-456, October, 1998.

[12] Ghosh, A., Schwartzbard, A., and Shatz, M., "Learning Program Behavior Profiles for Intrusion Detection," in Proceedings First USENIX Workshop on Intrusion Detection and Network Monitoring, Santa Clara, California, April 1999.

[13] Lippmann R.P., Cunningham, R.K., Improving Intrusion Detection Performance using Keyword Selection and Neural Networks, RAID Proc., West Lafayette, Indiana, Sep. 1999.

[14] Ryan, J., Lin, M., and Mikkulainen, R., "Intrusion Detection with Neural Networks," Advances in Neural Information Processing Systems, Vol. 10, MIT Press, 1998.

[15] Yang, J.H., Huang, S-H.S., and Wan, M.D., "A Clustering-Partitioning Algorithm to Find TCP Packet Round-Trip Time for Intrusion Detection," Proceedings of International Conference on Advanced Information Networking and Applications (AINA 2006), Vienna, Austria, Vol. 1, pp 231-236, April 2006.

[16] Wu, H.C., Huang, S-H.S., "Performance of Neural Networks in Stepping-Stone Intrusion Detection," IEEE International Conference on Networking, Sensing and Control (ICNSC), Sanya, Hainan Island, China, April, 2008 (accepted).
[17] Li, Q., Mills, D.L., "On the Long-range Dependence of Packet Round-trip Delays in Internet," Proceedings of International Conference on Communications, Atlanta, GA, Vol. 1. pp. 1185-1192, June 1998.
[18] Yang, J.H., and Huang, S-H.S., "Matching TCP Packets and Its Application to the Detection of Long Connection Chains," Proceedings of International Conference on Advanced Information Networking and Applications (AINA 2005), Taipei, Taiwan, pp 1005-1010, March 2005.
[19] Yang, J.H. "Detecting and Preventing Stepping-Stone Intrusion by Monitoring Network Packets," Ph D Dissertation, Department of Computer Science, University of Houston, Houston, Texas, August 2006.
[20] Russell, S. and Norvig, P., Artifitial Intelligence: A Modern Approach, 2nd edition, Prentice-Hall, 2003.
[21] NeuralWare, NeuraWorks Professional II Plus manual, Pittsburgh, PA, 2003.

Packet Fluctuation Approach for Stepping-Stone Detection

Han-Ching Wu and Shou-Hsuan Stephen Huang
Department of Computer Science, University of Houston, Houston, TX 77204, USA
Email: {hcwu, shuang}@cs.uh.edu

***Abstract* - To evade detection, network attackers usually launch intrusions through stepping-stones by building a long connection via intermediary hosts. In order to detect long connection chains, we first need to identify whether a host has been used as a stepping-stones. In this paper, we proposed the packet fluctuation approach to detect stepping-stones based on the range of a random walk model. Two algorithms (transformation and packet size) are proposed for this approach to distinguish the stepping-stone connections (*ATTACK* pairs) from the normal connections (*NORMAL* pairs). We also show the effectiveness of our algorithms in detecting the chaff perturbation. It is found that both algorithms are able to effectively identify the stepping-stone connections.**

Keywords: stepping-stone detection, random walk, packet fluctuation, chaff perturbation.

I. INTRODUCTION

Network intruders often launch attacks by constructing a long connection chain via intermediate hosts, called stepping-stones, in order to evade detection. There is no valid reason for a user to use a long connection chain to log into a host. Thus one way to prevent intruders is to detect long connection chains. There have been some studies done in this area including methods to estimate the length of the connection chains [1, 2, 3].

One way to decide if a host is being used as a stepping-stone is to find a connection pair by monitoring and comparing the incoming and outgoing connections of a host. An earlier method on stepping-stone detection with a content-based algorithm was proposed by Staniford-Chen and Heberlein [4]. They used thumbprints to compare streams, but thumbprints cannot be used for an encrypted session. Zhang and Paxson [5] proposed a method for encrypted connection chains to investigate correlation of connections based on stream timings and to match the on- and off-periods of connections. A deviation-based algorithm was reported by Yoda and Etoh [6], who used sequence number and packet delay to compute the deviation between traces. Donoho et al. [7] applied wavelet transforms to examine the maximum tolerable delay but they did not provide a bound on the number of packets. On the other hand, a watermark-based approach, proposed by Wang and Reeves [8], studied the independent and identically distributed perturbation for packet inter-arrivals. He and Tong [9] suggested two activity-based algorithms with bounded memory or bounded delay perturbation.

Blum et al. [10] proposed algorithms for stepping-stone detection using concepts from computational learning and random walks theories. Theoretically, they guaranteed the monitoring time and false positive rate without chaff perturbation. They also constructed an algorithm dealing with chaff that allowed a limited number of chaff packets by increasing the detection threshold. However, the number of packets needed is very large, and the allowed chaff rate for attacker to evade detection is too small to be practical.

In this paper, we will introduce the Packet Fluctuation approach to improve the method [10] mentioned above. Two algorithms will be proposed, namely *Transformation* and *Packet Size*. Based on the previous work [11, 12], the transformation algorithm applied a coordinate transformation to adjust for the difference in the numbers of packets in two streams. It resulted in a smaller bound range to detect the *ATTACK* pairs, and lowered the probability of false negative detection. In this paper, we will set the *chaff rate* as a parameter to investigate how different chaff rates can affect the algorithm ability to correctly identify *ATTACK* pairs. In Section IV, we will demonstrate another algorithm to take advantage of the packet size for stepping-stones detection. Based on the fluctuation caused by the size difference, we find a new way to compute the average score. It is found that the *ATTACK* pairs have the smaller average scores for the cases regardless the presence of the chaff perturbation. The experiments and results show that both methods can identify the stepping-stone pairs even for the chaffed streams.

This paper is structured as follows: Section II will discuss the fundamental background; Section III will investigate the Transformation method and analyze the experimental results; Section IV will describe the Packet Size method and discuss the experimental results; finally, Section V will conclude the paper with a brief summary.

II. FUNDAMENTALS

A. Background

The connection is represented by streams of packets. In general, for every packet sent in the incoming stream *S1*, there must be a corresponding packet in the outgoing stream *S2* within a delay bounded by Δ (a time interval), and let P_Δ represent the maximum number of packets transmitted in Δ. We define P_Δ the *maximum tolerable delay*. The number of packets sent in stream *S1* should be closely related to that in stream *S2* if the two connections are relayed. We recorded the correlation of the outgoing stream *S2* to the incoming stream

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 364–369.

S1 in the *S2-S1* versus *S2+S1* space, where *S2-S1* represents the accumulated difference in packet count and *S2+S1* (the accumulated sum) represents time steps [11]. We use *Z* ("*packet difference*") to denote *S2-S1* in algorithm and analysis.

Basically, if the two *connections* are relayed and not manipulated, the *Z* should fluctuate around zero. The reason that Z may not be exactly zero is that some packets may be split/combined during propagation, and some resent packets might present. However, for relayed connections, there is a statistically significantly higher probability that Z is around zero. In addition, if two relayed streams are manipulated, *Z* should be bounded within a range $[-P_\Delta, P_\Delta]$ under the assumption of maximum bounded delay. For example, in chaff perturbation, at most *r* superfluous packets are introduced in a unit time for each connection. If we collect the packets in *t* units of time, the *Z* should be within a range $[-P_\Delta, P_\Delta]$ for two relayed streams, where $P_\Delta = r*t$. Therefore, detecting a stepping-stone pair can be considered as examining if the differences of the number of packets between two streams are bounded (i.e. $-P_\Delta < Z < P_\Delta$). We call a pair of incoming and outgoing streams a *stepping-stone* pair or *ATTACK* pair if it is part of a stepping-stone attack. Otherwise, it is called *non-stepping-stone pair*, or a *NORMAL* pair.

The packet difference *Z* between stream *S1* and stream *S2* can be modeled by a one-dimensional random walk [8]. We assume that for each packet on the union of the two streams, it comes from *S1* with probability *p* and from *S2* with probability *q*. In other words, every time a packet is sent on either stream *S1* or *S2*, *Z* decreases by 1 and walks a step toward negative direction with probability *p*, and increases by 1 toward positive direction with probability *q*.

Assume the random walk starts at the origin. If the two streams belong to a stepping-stone pair, then *Z* should be bounded within a range $[-P_\Delta, P_\Delta]$. Otherwise, *Z* will exit the range in time [13]. The question is when *Z* will exit the range $[-P_\Delta, P_\Delta]$. Based on the result of the random walk process [13]: if $p = q = ½$, the expected time to exit the range from either boundary is $(P_\Delta+1)^2$. If the connection chain is manipulated, such as chaff perturbation, the probability that a packet might be in stream *S1* with *p*, and in stream *S2* with *q*, where we are not sure which one is larger, i.e. $p \neq q$. For example, if more packets were inserted into stream *S1*, *p* would be greater than *q* and *Z* would walk toward the negative direction; otherwise, it would walk toward the positive direction.

B. Chaff Rate

Blum et al. [10], first proposed the detection algorithm without chaff, DETECT-ATTACKS (DA), for a given pair of streams. They derived the total monitoring number of packets required to be $2(P_\Delta+1)^2$. To evade detection, the attacker could add chaffed packets into streams *S1*, *S2* or both. This chaff perturbation is not necessary to reach the destination. Normally, the amount of chaff is under a predefined limit. Blum et al. [10] developed another algorithm, DETECT-ATTACK-CHAFF (DAC), to detect stepping-stone pairs with chaff. However, DAC could not detect the attacking connection if an attacker inserts sufficient chaff and causes the *Z* between the two streams to be greater than $2P_\Delta$ packets; i.e. the attacker can evade the detection by sending more than P_Δ packets of chaff in every time interval.

In [10], P_Δ is used as the maximum bound for the chaff packets. In this condition, the DAC algorithm can detect stepping-stone attacks only under a very limited amount of chaff that is less than $P_\Delta+1$ packets of chaff for every $8(P_\Delta+1)^2$ packets. The chaff rate is less than $1/(8P_\Delta)$. As shown in Table 1, the chaff rates used in Blum et al. [10] under confidence level (δ) of 10% are too small. For example, if P_Δ is from 1 to 100, the chaff rates can only reach 1.88% to 0.08%, respectively, as shown in Table 1 below.

Table 1. Chaff rates in Blum et al. [10] (δ=10%)

P_Δ	1	10	50	100
M_B	106	3215	69122	271095
chaff rate	1.88%	0.62%	0.14%	0.08%

Since P_Δ and the maximum number of chaff packets should be two unrelated quantities, in our paper, we propose to study the effect of "*chaff rate*" rather than using a fixed amount of chaff packets. The chaff rate is defined as the ratio of the number of chaff packets to the minimum of packet numbers in streams *S1* and *S2*. We can use *p* (number of packets of incoming stream) and *q* (number of packets of outgoing stream) to represent the chaff rate as:

$$Chaff\ rate = \frac{|p-q|}{((p+q)-|p-q|)/2} \cdot 100\% \quad (1)$$

In the following section, we will go over a new transformation algorithm to decrease the packet bound in order to detect the stepping-stone connections. Furthermore, we conducted some experiments to demonstrate the effect of the improved algorithm.

III. TRANSFORMATION METHOD

A. Algorithm

If we compute a linear regression of the packet difference *Z*, the line will have a positive or negative slope. For example, assume streams *S1* and *S2* are an attacking pair. If *S1* is chaffed, it will result in a negative slope of the curve, as shown in curve (2) of Figure 1. Thus given enough time, *Z* will exit any given bound. In other words, if we collect enough packets, there will be false negative cases.

The false negative issue does not conflict with Blum's claim of no false negative cases because in DAC algorithm the assumption was a fixed maximum chaff number. Thus the more packets were collected, the lower the chaff rate would be. When P_Δ was a fairly small number (say, 10), the DAC algorithm performed very limited use, for the intruders might insert much more packets during the whole session. Therefore, in contrast to Blum's fixed maximum chaff number, we hypothesize that using a fixed chaff rate will be a better and more realistic parameter.

Constant chaff rate might result in an enormous quantity of the chaff packets. For example, as shown in curve (1) of Figure 1, adding 200 chaff packets into *S1* brought the chaff rate up to 102%. Due to the negative slope, the *Z* would easily exit the given bound, resulting in false negative cases. However, we reasoned out to use the "*fluctuation*" of *Z* to detect the attacking pair, rather than the "*slope*" of the curve. This logically led us to apply the transformation method for stepping-stone detection. The transformation made the slope in Figure 1 equal to zero. Consequently, the fluctuation would be around zero, canceling the effect of the chaff.

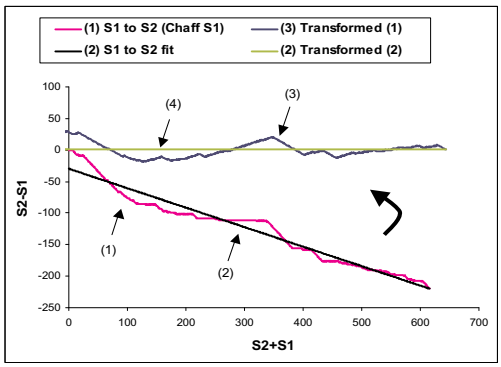


Figure 1. Correlation curves vs. corresponding linear fits before and after transformation

To demonstrate the effect of the improved algorithm, we conducted several experiments. The goal was to distinguish stepping-stone connection from the other non-stepping-stone connections. We performed tests on the connections of incoming stream *S1* and outgoing stream *S2*, and recorded packets in those streams. At the end of the experiment, our program would relate stream *S1* to stream *S2* in the space of *S2–S1* versus *S2+S1*, as shown in Figure 1. It should be noted that the linear fits are also plotted for both before and after transformation.

In the following, we present an improved method to take into account the issue of non-zero slope. By applying a simple transformation, we forced *Z* to center around the x-axis, thus eliminating the effect of the non-one-to-one problem.

According to Blum [10], the two primary parameters in the algorithms are the false positive rate δ and the boundary P_Δ. If P_Δ is unknown, P_Δ is assigned to the maximum value of the difference between the numbers of packets generated by the two streams. The maximum and minimum values in the magnitudes of *S2–S1* along the curve correspond to the upper and lower bounds, respectively, and the boundary range is calculated by the difference of these two bounds.

If a curve had a non-zero slope, the magnitude of *Z* would grow with increasing the time. Hence, for a given P_Δ, the magnitude of *Z* would easily exit P_Δ and fail to find true attacking connection, resulting in a false negative detection. It was noted that the experimental curve before transformation showed a large boundary range. To shorten the boundary range of the curve, we performed a coordinate transformation for all points on the curve. After transformation, the points of the new curve were distributed around the x-axis. The experimental description and results of both with and without chaff were reported in [12].

B. Effect of Chaff Rate

We were interested in finding out the effect of the chaff rate on the ability for us to identify the correct pairing of the streams. In the following experiment, we set up the incoming and outgoing streams on a host in our lab: *IN1*, *IN2* and *IN3* stood for the three incoming streams; *OUT1*, *OUT2* and *OUT3* represented the three outgoing streams. We established three chains to off campus locations on the Internet. We started the three connection chains at the same time with one of the incoming stream matching the outgoing stream. On the outgoing stream, we added chaff packets at different rate. These chaff packets were generated using a Poisson distribution and added to the packets collected from the experiment. This allowed us to isolate the effect of the chaff rate. Table 2 shows the experimental results of three test cases *c1, c2* and *c3*, in which only the outgoing stream was chaffed.

Table 2. Experimental results with chaff rate = 100%

		Before Transformation			After Transformation		
	Connect	LB	UB	Range	LB	UB	Range
c1	**IN1 to OUT1***	**-1**	**293**	**294**	**-39**	**18**	**57**
	IN2 to OUT1*	0	381	381	-76	98	174
	IN3 to OUT1*	-38	119	157	-52	86	138
c2	IN1 to OUT2*	-20	260	280	-144	62	206
	IN2 to OUT2*	**-1**	**197**	**198**	**-30**	**18**	**48**
	IN3 to OUT2*	-78	71	149	-104	56	160
c3	IN1 to OUT3*	0	629	629	-77	41	118
	IN2 to OUT3*	0	712	712	-54	65	119
	IN3 to OUT3*	**-1**	**454**	**455**	**-36**	**23**	**59**

* chaffed stream; LB: lower bound; UB: upper bound

In each set of experiment, we selected each *OUT* to match *IN1, IN2* and *IN3*, thus obtaining three sets of data – which all contained two *NORMAL* pairs and one *ATTACK* pair of streams. For each pair, we could compute the range of the random walk. Typically, the *ATTACK* pair had the smallest range. However, to correctly identify the *ATTACK* pair, we would like to see the range of the *ATTACK* pair to be significantly smaller than that of the *NORMAL* pairs. As a result, in Table 2 we took the smallest *NORMAL* pair ranges (138, 160 and 118) minus the *ATTACK* pair ranges (57, 48 and 59), respectively. The minimum margin (minimum of 81, 112 and 59) could be used as a measure to see how the algorithm worked for different chaff rates. In Table 2, the minimum margin of 59 was large enough for us to identify the *ATTACK* pairs for each example.

Figure 2 demonstrated the relationship between the different chaff rates and the minimum margin (i.e. algorithm effect to correctly identify the *ATTACK* pairs). Chaff rates from 0% to 600% with an increment of 100% each time were tested. As

the chaff rate increased, the minimum margin dropped as predicted. When the chaff rate reached 300% and beyond, the minimum margin became so small and trivial that it was difficult to identify the *ATTACK* pairs of streams. Furthermore, at the 600% chaff rate, the algorithm failed to identify the *ATTACK* pair (i.e. minimum margin was recorded below zero). This result described above was, however, an improvement from the range detection without transformation in which the algorithm failed at 50% chaff rate (not shown in table/figure).

Our experiments suggested that we could use a threshold of 100 as an absolute cutoff point to identify pairs. The number worked on our data up to 200% chaff rate. However, the number of experiments we performed was limited, so further test data need to be collected.

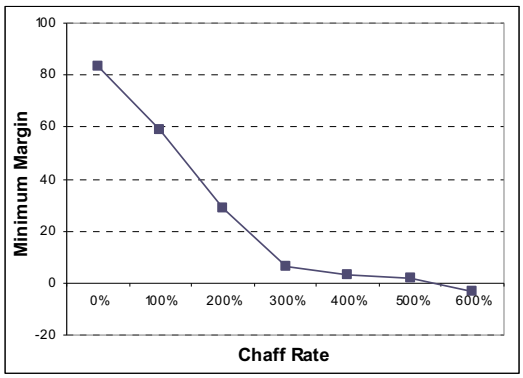


Figure 2: Minimum Margin vs. Chaff Rate

IV. PACKET SIZE METHOD

For an encrypted connection, all packet contents are encrypted, so they cannot be used to make comparisons. The only information that can be used is from packet header with timestamp, size and other TCP protocol. Throughout the rest of this paper, we will take advantage of packet sizes and propose the *Packet Size* method to detect the stepping-stone pairs.

We collected the packet sizes in each of the incoming stream *S1* and the outgoing stream *S2*. Because the packet of *S1* was from previous host whose packet size might be different from that of the monitoring host. Even for the relayed connection pair, both packet sizes might be different. We observed the correlation of stream *S2* to stream *S1* in the *Size2-Size1* versus *S2+S1* space, where *S2+S1* was the accumulated sum of the numbers of packets indicating time step (same as the previous section). *Size2-Size1* was the accumulated difference of the packet sizes in *S1* and *S2*. We use *W* ("*size difference*") to indicate *Size2-Size1* in the algorithm.

As shown in Figure 3, for the relayed connection (*IN1* to *OUT1*), the fluctuation caused by size difference was very regular because most sent packets contained only one character. In contrast, the un-relayed connection (*IN1* to *OUT2*) showed much larger and irregular fluctuation in a greater magnitude. Without using transformation, we proposed another packet size algorithm to detect the stepping-stone pair.

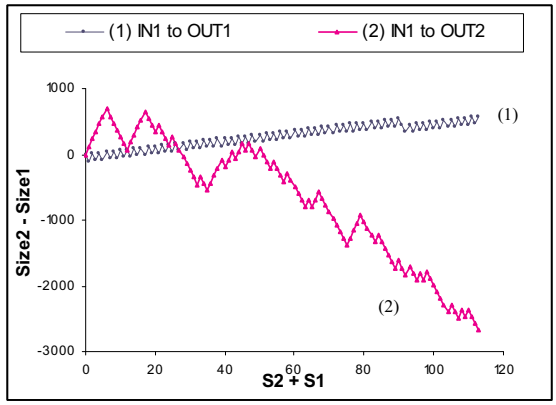


Figure 3. Comparison of *ATTACK* and *NORMAL* pairs using accumulated size difference

Algorithm *A* (Figure 4) was used to find the difference of the accumulated stream sizes. Another algorithm *S* (Figure 5) served to compute the average score in order to investigate the fluctuation magnitude of a given pair of the connection streams. In other words, the average score was the average

Algorithm A:

Input:
- One incoming stream *IN[1:n]* in ascending order;
- One outgoing stream *OUT[1:m]* in ascending order;

Output:
- Accumulated streams *S2+S1, Size2-Size1*;

```
Initialize S1,S2; //accumulated counts of IN & OUT streams
Initialize Size1,Size2; //accumulated sizes of IN & OUT
i = j = k =1;
while(i<=n && j<=m) {
   if(IN[i].time < OUT[j].time){
      S1[k] = S1[k-1] + 1;
      S2[k] = S2[k-1];
      Size1[k] = Size1[k-1] + IN[i].size;
      Size2[k] = Size2[k-1];
      i++; }
   else {
      S1[k] = S1[k-1];
      S2[k] = S2[k-1] + 1;
      Size1[k] = Size1[k-1];
      Size2[k] = Size2[k-1] + OUT[j].size;
      j++; }
   k++;
}
While (n>m) add remaining IN to S1, Size1;
While (n<m) add remaining OUT to S2, Size2;

Return (S2+S1, Size2-Size1);
```

Figure 4. Algorithm *A* to compute *S2+S1* and *Size2-Size1*

packet size fluctuation. The average score was used instead of the total score, eliminating the difference of the accumulated packet numbers in different connection pairs. In the paper, we use the score to represent the average score.

Algorithm S:

Input:
- Size difference array W[1:k] //W = Size2-Size1, k=n+m;

Output:
- Score_Avg //average score for the fluctuation magnitude

```
for (i=1 to k-1)  F[i] = W[i] – W[i+1];
for (i=1 to k-1) {
   if (F[i] > 0 and F[i+1] < 0)
       or (F[i] < 0 and F[i+1] > 0)
       //fluctuation in different direction
      score[i] = 0;
   else
      score[i] = 1; //in same direction
   Score_Sum += score[i];
}
Score_Avg = Score_Sum/k; //average score
Return (Score_Avg);
```

Figure 5. Algorithm *S* to compute the average score

We conducted the first experiments on the connections of three incoming streams and three outgoing streams. Table 3 shows the results from algorithms *A* and *S* for the case without chaff. Obviously, the *ATTACK* pairs along the diagonal had much smaller score values. The remaining six scores of *NORMAL* pairs were much larger. Like the previous method, the algorithm could distinguish *ATTACK* pairs from *NORMAL* pairs without chaff

Table 3. Experimental results without chaff

	IN1	IN2	IN3
OUT1	**0.0150**	0.8931	0.6346
OUT2	0.9002	**0.0319**	0.8450
OUT3	0.6427	0.8448	**0.0171**

In our second experiments, we took into account the chaff packets. Table 4(a) and 4(b) shows the experimental results by adding 100% rate of chaff packets on the outgoing stream *S2* and the incoming stream *S1*, respectively. Although the chaff perturbation increased the scores of the *ATTACK* pairs, the scores along the diagonal were still smaller than the other six scores of *NORMAL* pairs. Our experimental data showed that this packet size algorithm could distinguish the *ATTACK* pairs from the *NORMAL*.

Similar to the transformation method, we would like to find the largest chaff rate that was able to identify the ATTACK pairs. Finding the minimum margin was the way to obtain the score range of the ATTACK pair that was significantly smaller than that of the NORMAL pairs. For example, in Table 4(a) we took the smallest NORMAL pair scores in each row (0.6026, 0.7013 and 0.6778) minus the ATTACK pair scores (0.3374, 0.3483 and 0.3331), respectively. The minimum margin (minimum of 0.2652, 0.3530 and 0.3447) could be used as a measure to see how the packet size algorithm worked for different chaff rates. In Table 4(a), the minimum margin 0.2652 was sufficiently large to identify the ATTACK pairs. Similar procedures were done for Table 4(b).

Table 4(a). Experimental results with chaffed outgoing streams (chaff rate = 100%)

	IN1	IN2	IN3
OUT1*	**0.3374**	0.8835	0.6026
OUT2*	0.7977	**0.3483**	0.7013
OUT3*	0.6778	0.8680	**0.3331**

* chaffed streams

Table 4(b). Experimental results with chaffed incoming streams (chaff rate = 100%)

	IN1*	IN2*	IN3*
OUT1	**0.3463**	0.8132	0.6728
OUT2	0.8892	**0.3599**	0.8626
OUT3	0.5966	0.7047	**0.3478**

* chaffed streams

Table 5 Minimum margins for average scores with different chaff rates

Chaff Rates	0%	50%	100%	200%	300%
IN*	0.6175	0.3903	0.2503	0.1108	0.0297
OUT*	0.6196	0.3866	0.2652	0.0957	0.0416

* chaffed streams

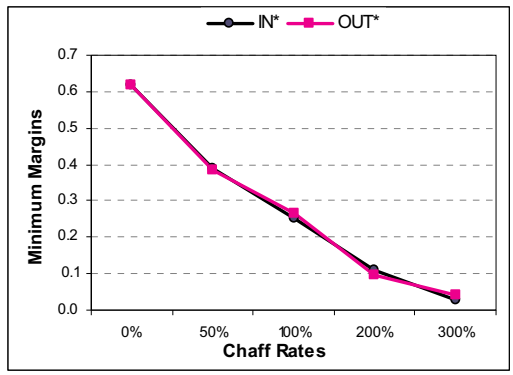


Fig. 6: Minimum Margins of Fluctuation vs. Chaff Rates

Table 5 and Figure 6 demonstrated the relationship between the different chaff rates and the minimum margin. Chaff rates from 0% to 300% with an increment of 100% each time (except 50% chaff rate) were tested. The results from both chaffed incoming and outgoing streams were very close. As the chaff rate increased, the minimum margin dropped as predicted. When the chaff rate reached 300%, the minimum margin was very close to zero and was hard to identify the

ATTACK pairs. We could apply a threshold of 100% as an absolute cutoff point to identify pairs. From the experimental results, the proposed packet size method could be applied to detect the stepping-stones for the cases with and without chaff perturbation.

The results shown in this paper focused on the Send packets. Our future study will concentrate on the Echo packets, in which we expect to obtain better results because the Echo packets usually contain more large packets that can serve as the markers to detect the stepping stones.

V. CONCLUSIONS

In this paper, we have described our recent studies in stepping-stone detection. First, we propose an approach to reduce the packet bound by performing a coordinate transformation of the packet difference of two streams to distinguish *ATTACK* connection from the other *NORMAL* connections. Instead of using a fixed limit on the total chaff packets proposed by Blum et al. [10], we set a chaff rate as a parameter, which might be closer to a real attack. It was found that after the transformation, the boundary range was much smaller for *ATTACK* connection, resulting in smaller probability of false negative detection. The smaller bounds also mean fewer packets to be examined in order to gain enough confidence according to random walk. In the second approach, we developed a packet size method, in which the packet size in both incoming and outgoing streams were collected. From the fluctuation caused by the accumulated size difference, we computed the average scores to compare the attacking and normal connection pairs. The results showed that the *ATTACK* pairs had smaller scores and the *ATTACK* pairs could be identified for both with and without chaff perturbation. The results presented in this paper were based on the Send packets. In the future, the Echo packets will be studied. Since the Echo packets generally contain more large packets, it is believed that better results will be obtained.

ACKNOWLEDGEMENT

This project is supported in part by a grant from Texas Learning and Computation Center (TLC^2).

REFERENCES

[1] Yung, K. H., "Detecting Long Connection Chains of Interactive Terminal Sessions," In Recent Advances in Intrusion Detection (RAID), LNCS 2516, pp. 1-16, January 2002.

[2] Yang, J. and Huang, S.-H. S., "A Real-Time Algorithm to Detect Long Connection Chains of Interactive Terminal Sessions," Proceedings of the 3rd international conference on Information security. New York, NY, USA: ACM Press, pp. 198-203, 2004.

[3] Yang, J. and Huang, S.-H. S., "Matching TCP Packets and its Application to the Detection of Long Connection Chains on the Internet," Proceedings of the 19th International Conference on Advanced Information Networking and Applications, Taipei, Taiwan, pp. 1005-1010, March 2005.

[4] Staniford-Chen, S. and Heberlein, L.T., "Holding Intruders Accountable on the Internet", in Proc. of the 1995 IEEE Symposium on Security and Privacy, Oakland, CA, pp. 39-49, May 1995.

[5] Zhang, Y. and Paxson, V., "Detecting Stepping Stones", in Proc. of the 9th USENIX Security Symposium, Denver, CO, pp. 171-184, August 2000.

[6] Yoda, K. and Etoh, H., "Finding a Connection Chain for Tracing Intruders", 6th European Symposium on Research in Computer Security-ESORICS 2000 LNCS-1895, Toulouse, France, pp. 31-42, October 2000.

[7] Donoho, D., Flesia, A.G., Shankar, U., Paxson, V., Coit, J. and Staniford, S., "Multiscale Stepping-Stone Detection: Detecting Pairs of Jittered Interactive Streams by Exploiting Maximum Tolerable Delay", 5th International Symposium on Recent Advances in Intrusion Detection, Lecture Notes in Computer Science 2516, New York, Springer, 2002.

[8] Wang, X. and Reeves, D., "Robust Correlation of Encrypted Attack Traffic Through Stepping Stones by Manipulation of Inter-packet Delays", in Proc. of the 2003 ACM Conference on Computer and Communications Security (CCS 2003), ACM Press, pp. 20-29, October 2003.

[9] He, T. and Tong, L., "Detecting Encrypted Stepping-stone Connections", IEEE Trans. on Signal Processing, Vol. 55, No. 5, pp. 1612-1623 May, 2007.

[10] Blum A., Song D. and Venkataraman S., "Detection of Interactive Stepping Stones: Algorithms and Confidence Bounds," in Proc. of 7th International Symposium on Recent Advances in Intrusion Detection (RAID '04), Springer LNCS 3224, pp. 258-277, 2004.

[11] Huang, S-H. S., Lychev, R. and Yang, J., "Stepping-Stone Detection via Request-Response Traffic Analysis", Proceedings of 4th International Conference ATC, LNCS 46100, pp. 276-285, 2007.

[12] Wu, H.C., Huang, S-H.S., "Detecting Stepping-Stone with Chaff Perturbation," Proceedings of AINA, International Symposium on Frontiers in Networking with Applications (FINA), Ontario, Canada, Vol. 1 pp. 85-90, May 2007.

[13] Feller, W., "An Introduction to Probability Theory and its Applications", Volume 1. John Wiley and Sons, Inc., 1968.

Using Mobile Telephone as an Operator Independent, Secure Micro-Payment Tool

Hasan AMCA[*] and Erbug CELEBİ[**]
[*]Electrical and Electronic Engineering Department, Eastern Mediterranean University, hasan.amca@emu.edu.tr
[**]Computer Engineering Department, Cyprus International University, ecelebi@ciu.edu.tr

ABSTRACT

With the latest advances in Information and Communication Technologies (ICT), using different technologies for electronic-payment has become a major issue in the retail market. The use of portable communication devices became particularly attractive candidates when versatility, security and simplicity features of payment technologies are considered.

In this paper, we investigate the use of mobile communication devices as versatile, secure and simple micro-payment tools, which satisfy the related financial, technological, computational and managerial requirements. The versatility and security of the method comes from the use of a mobile telephone and a Variable Transaction Number (VTN) in each transaction. Experimental results have shown that, the systematic requirements for the implementation of this technology are minimal and the costs involved are very much reasonable.

1. INTRODUCTION

Due to the fraudulent use, loss or damage of the card-based electronic payment (e-payment) devices such as Visa, Master-Card, Card Plus and American Express, there is a significant annual financial loss [1]. In addition, the implementation of electronic signature introduced additional security problems due to the lack of facilities to enter the Personal Identification Number (PIN) to the Point of Sale (POS) terminals. A simple but yet versatile and secure electronic payment technology could be implemented by the use of Mobile Telephones (MT), that will eliminate the security related problems due to the use of card based electronic payment devices.

MTs could be used in e-payment in several different ways, such as, SMS, IrDa, Bluetooth, RFID [3, 16]. These methods have the common ground of charging the mobile telephone for the purchases made [4] and integrate the purchasing expenses and mobile phone bill.

SMS method allows users to anonymously and securely pay for the products and services they purchased via their mobile phone by sending a text message to a premium number. The customers are then charged on their mobile phone invoice. The SMS method is designed to work in batch processing mode and therefore might take a long time to confirm credit approval by the bank and complete the transaction [5].

Despite the well defined IrDa specifications, the relatively long setup time renders IrDa useless for m-payment. The "express payment", designed to reduce the setup time, reduces the transaction time significantly. However, it also reduces the security level by giving the privilege to devices to bill the consumer without authentication and making them potentially vulnerable to financial fraud [2,6].

Bluetooth requires a relatively long setup time before the payment process starts. Plus, it has a non-selective nature. A Bluetooth device will search for all devices within a short range. This might mean a large number of Bluetooth devices in a shopping mall [7]. The usage of RFID in m-payment could be comparable to Bluetooth and therefore is delay and complexity limited [8,9].

The method we proposed in this article uses a credit provider generated secure transaction number, unique for each transaction. This number is transmitted through the GPRS channel and transformed into a barcode by the MT that can easily be read by the merchant's barcode reader. Hence, overcoming the shortcomings mentioned above in terms of latency, security and usability.

This method of payment could be made available in societies where Mobile Telephone usage is more widespread than the Credit Card (CC). Most of the underdeveloped and developing countries have higher rate of mobile telephone penetration than CC. Hence, m-payment in such countries can help improve spreading of electronic payment technologies by using MTs instead of CCs.

This article is organized as follows: In section 1, an introduction is made into the m-payment technologies. In section 2, a preview of online payment processing is presented. In section 3, the Variable Transaction Number Barcode (VTNB) method is described with reference to the infrastructure and security issues. Experimental setup and test results are presented in section 4. Conclusions are provided in the last section.

2. ONLINE PAYMENT PROCESSING METHODS

Electronic payment could be referred to as payment for the purchase of goods or services without using hard cash. An electronic payment solution should be versatile, secure and easy to use so that the common fraud-related risks such as product theft, identity theft and cash theft will be avoided. To satisfy these requirements, an electronic payment solution should consists of 2 steps: *Authorization* and *settlement*.

Authorization verifies that the payment tool is active and the customer has sufficient credit to make the transaction. This is shown in Fig. 1. *Settlement* is the process of charging the customer's payment account and transferring money from the customer's account to the merchant's account through a

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 370–374.

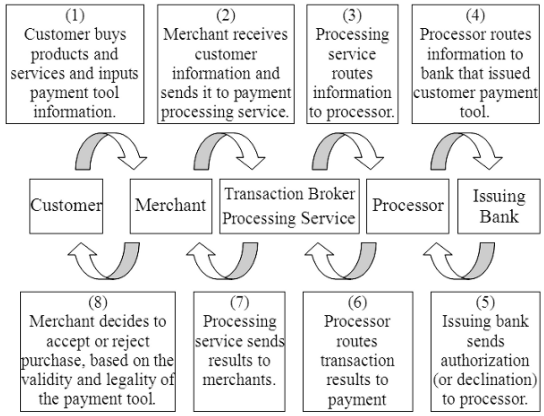


Fig. 1: Payment Processing Authorization cycle.

transaction broker such as PayPal [14]. This is depicted in Fig. 2 below. The following sections describe the theoretical background and practical implementation of an electronic payment system using a MT as a payment tool. The authentication and the settlement cycles are also displayed.

3. VARIABLE TRANSACTION NUMBER BARCODE (VTNB) METHOD

The repetitive use of the fixed credit cards and the fixed verification numbers in all transactions poses a major threat on the use of card based electronic payment system, since such numbers are easy to remember and relatively easy for some attackers to steal them. Some common ways of unlawfully obtaining credit card information such as shoulder surfing, dumpster diving, packet intercepting and database stealing are presented in [13]:

Due to the fraudulent use, loss or damage problems, yearly, there is a significant financial loss [2]. Not only does the CC fraud cause money loss, but also significant worry among customers. Hence, a successful method for payment should eliminate these problems and allow customers to use the

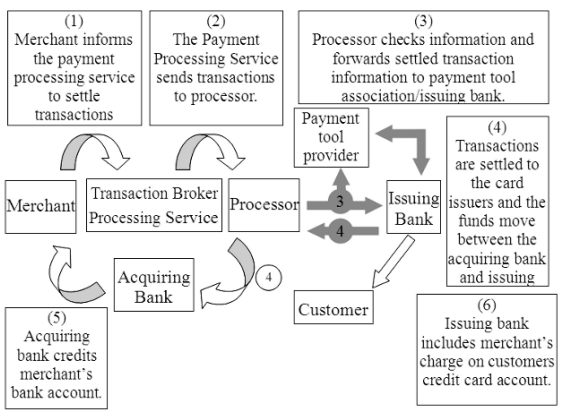


Fig. 2: Payment Processing Settlement cycle.

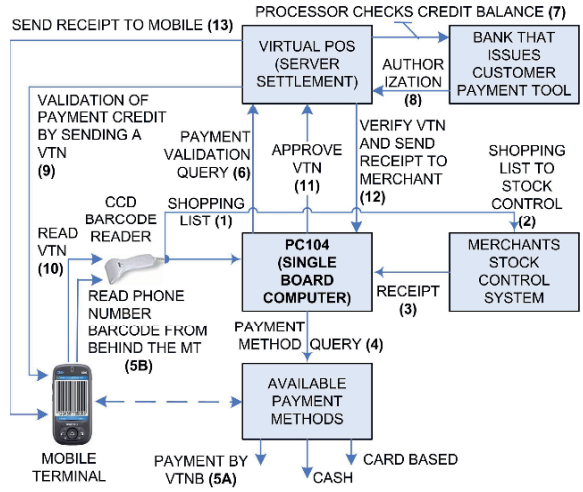


Fig. 3: VTNB Payment processing system authorization cycle.

payment technique without worry.

As a candidate for secure payment method, VTNB method should depend on different factors such as: systematic simplicity and feasibility, open architecture, short transaction delay, ease of use by the target customers, interoperability between different vendors and security. The VTNB method will use the existing telecommunications infrastructure and satisfy all of the above criteria as follows.

3.1. Systematic Simplicity and Feasibility

The systematic simplicity and feasibility refers to the additional hardware and software required to build the VTNB over the existing payment system. The VTNB system should also be fast, traversing minimum number of proprietary networks. This way, payment of small amounts is possible since network usage is limited and the overhead is low. This can be better understood with reference to Fig. 3 and Fig. 4. The VTNB method of payment is basically similar to the credit card method except that, in the former the onscreen barcode is read by the barcode reader and manipulated by the terminal device replacing the POS terminal. The terminal device is a single board computer with a GPRS support and a USB sockets. A PC104 running on

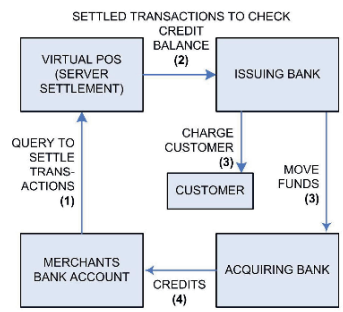


Fig. 4: VTNB payment processing system settlement cycle.

Fig. 5: A PC104 single board computer running on Linux used to produce the terminal device.

Linux could be used to produce the terminal device as shown in Fig. 5.

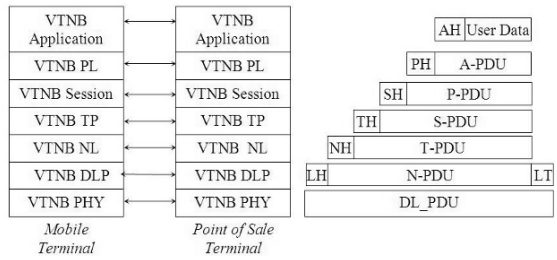
3.2. Layered Protocol Architecture

The widespread acceptance of the VTNB method depends on the choice of the architecture and the communication protocol used. The architecture should support the Open System Interconnection (OSI) architecture so that different vendors could produce electronic equipment such as the VTNB processor compatible with the barcode reader and the merchants stock control system. The layers of VTNB protocol stack are shown in Fig. 6 [10, 11]:

The application layer of VTNB hosts the application program responsible for receiving the VTN through the GPRS channel and performing the necessary operations. Steps 5, 9, 10 and 13 in Fig. 3 are defined in the VTNB Application Layer Protocol and are described in the following steps.

After the transaction amount is displayed on the merchant's terminal screen, the customer decides to pay by VTNB method. Payment is initiated by reading the MT phone number in the form of barcode from the back cover of the MT (5B). The phone number is sent to the VPOS through wireline network and the VPOS sends the VTN to the MT if the phone number is authorized for such payment. The VTNB is presented to the merchant's Charge Coupled Device (CCD) barcode reader to finalize the payment process (steps 5 to 10) is followed by verification of the VTN by the VPOS (12). The on-line connection to the VPOS is provided by the module on the cash register instead of the POS device.

An additional security level is created by introducing the entry of PIN number to the MT during the payment cycle right after step 7. Once the user accepts (by entering the PIN code on the MT) to pay the amount on the merchants screen, the payment process is completed. The VTNB payment system is suitable for other applications such as petrol stations, vending machines, buses, car parks, ATM's, cinema and theatre entrance, classroom attendance check etc.



(A,P,S,T,L,N)H: (Application, Physical, Session, Transport, Network, Link) Header
PDU: Protocol Data Unit, LH: Link Header, LT: Link Trailer

Fig. 6: The 7 layer protocol stack for connection management between the MT and the POS terminal.

3.3. Transaction Delay

Transaction delay is one of the most important factors for acceptability of an electronic payment system. Literature survey has shown that, ideally, such a delay should be less than 1 second [12]. The transaction delay in VTNB system is expected to be shorter than that of the Bluetooth, RFID and IrDA but similar to that of the conventional Credit Card (CC) method. The delay is due to the reading of the phone number barcode, applying to the VPOS for a VTN, validation and granting VTN, reading VTN by barcode reader, verification of VTN and billing mechanism. The VTN is randomly drawn from a set with a negotiation between the financial service provider and the mobile terminal (MT), preferably as a function of the International Mobile Equipment Identity (IMEI) number of the MT. The provision of the VTN is the major challenge in the proposed method of payment and more research is to be made to eliminate delay in generating VTN without compromising security. The delays involved are presented in Fig. 7. Parameters effecting delay are; set-up time, connection time, data transfer time, processing time, security verification time and disconnection time. The following section clarifies the delay mechanism which starts after reading and transmitting the MT number barcode from the back cover of the mobile telephone (5B). The LOOP delay in VTNB Mobile Payment System includes,

T_1: Payment validation query transmitted to the VPOS (6),

T_2: Credit balance check and validation (7,8),

T_3: Validation of the payment credit by sending a VTN to the MT (9),

T_4: Reading the VTNB (10),

T_5: Sending the VTNB to the VPOS for approval (11),

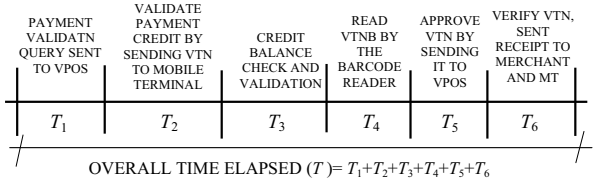


Fig. 7: The delays involved in VTNB method.

T_6: Verifying the VTN and sending the receipt to both the merchant and the customer (12,13).

3.4. Ease of Use by the Target Customer

The usage of the VTNB technology is as simple as sliding a good in front of the barcode reader, basically, nothing more than thinking the mobile telephone as a good. The users are not expected to have any special skills to adapt to the method. The only requirement is the need to register for the service prior to usage. In countries where the use of credit card is not widespread but the MT penetration rate is high, the VTNB system is expected to find high level of public acceptance.

3.5. Interoperability between Different Vendors

A set of global foundational m-payment standards need to be agreed upon in order for content providers to reach a critical mass of paying customers who, in turn, will then have a plentiful supply of applications and services to choose from. This will enable widespread availability of m-payment and the target customer range will also increase.

For the VTNB method to pick up easily and spread all around the world, a standardized communication and interconnection standard should be adapted in order for the equipment manufacturers to produce devices that will interoperate with the VTNB system. This requirement is satisfied by the use of an open architecture.

3.6. Security

VTN is made random in order to avoid reproduction without the MT holders consent. VTN should be accompanied by the PIN number of the MT user. The process of entering PIN number to the POS machine could easily be replaced by entering the PIN number on your mobile, in privacy and comfort of your palm. User portability is provided by the SIM card and the MT number barcode behind the telephone.

Other security issues: The loss or theft of consumers' wallets with their physical credit cards will not be noticeable until the next time they carry out a purchase. This can occur anytime from immediately to several days later. However, the awareness of a loss or theft of one's MT can be felt immediately.

Keeping and protecting the paper receipts for future reference is also a security issue in m-payment method. Paper receipts can be lost and cause consumer inconvenience and dissatisfaction in CC payment systems. However, the storage nature of the electronic receipts helps to protect them and work them out easily in the electronic form.

4. EXPERIMENTAL SETUP AND TEST RESULTS

The VTNB system, which is designed with an embedded system approach, is implemented on an x86 based Single Board Computer (SBC) running on Ubuntu Linux (specially designed for embedded systems) with low memory and CPU resources. In the embedded system approach, the system is designed to run for a relatively long time without errors. In the case of errors, a self recovery mechanism such as a watchdog timer is usually employed. No mechanical or moving parts such as hard disks are included. The embedded systems are less tolerant to errors. Therefore, they have a more sophisticated and thorough testing procedures.

The final system prototype is made up of, a 133MHz CPU, 64MB RAM, IDE interface, 2×USB ports, PCI interface, a parallel port, 2×RS232 ports, AGP compliant VGA interface, 44-pin IDE flash disk to store the system and application programs, keyboard, VGA display and +5V power supply. In addition, a Hitachi HD44780 based 20×4 characters LCD controller with an LCD4Linux driver as a display means.

This SBC will replace the POS terminal at the merchants counter. An application program using LibCurl[1] is developed and installed on the SBC as shown in Fig. 5. An application program that generates the VTN is also developed to run on the web server. In addition to VTN management, this application program will be responsible for bank transactions, keeping authorization logs, managing the database and generating records to enable system tracking. A Java application program that will aid receiving the VTN and converting it into a barcode is also developed using Midlets, for installation on the MT.

The preliminary test results have shown that the MT can be used as an electronic payment device along with the VTNB support framework. The loop delay is estimated through empirical measurements and timing calculations between the source and destination, resulting in T_1=0.2 sec, T_2=0.2 sec, T_3=2.0 sec, T_4=1.0 sec, T_5=0.2 sec, T_6=2.0 sec. The overall delay is therefore around 5.6 seconds, which is of the same order as the CC usage delay as our empirical study has shown. The security checks are fully operational and no run-time errors or defects is observed during 120 hour test run.

5. CONCLUSION

Card based electronic payment services have been widely used throughout the world in full confidence for several decades. However, due to the security gap in these services, alternative solutions have been sought for by the financial service providers. The use of mobile telephones appeared to be the closest competent to card based payment due to the advanced security features, reliability, interoperability between different vendors, ease of use by the target customers, high penetration rate, simplicity in implementation and technical feasibility. The system also guarantees that, the service provisioning by banks, operators and terminal manufacturers are also independent from each other.

The mobile telephones based VTNB system, which is introduced as an alternative electronic-payment scheme, has proven to be efficient, fast and highly secure. Through the use of PIN numbers for authentication, VTNB could yield an even higher level of security, The method could be easily customized to support either credit or debit based payment schemes.

[1] LibCurl is a common library used to grant web access to mobile devices.

6. REFERENCES

[1] Eliminating Some Credit Card Risk for E-Business, http://ecommerce.Internet.com/solutions/ec101/article/0,1467,6321_569741,00.html.

[2] Internet Usage Statistics – The Big Picture, http://www.internetworldstats.com/

[3] Pi Huang And A.C. Boucouvalas, Future Personal "E-Payment: Irfm", IEEE Wireless Communications, pp. 60-66, Feb. 2006.

[4] S. Schwiderski-Grosche and H. Knospe, "Secure Mobile Commerce", *Electronics & Communication Engineering Journal*, October 2002, pp. 228-238.

[5] S. F. Mjpllsnes and C. Rong, "On-line e-wallet system with decentralized credential keepers," *Mobile Network Applications,* vol. *8 ,* pp. 87--99,2003.

[6] IrDA, Infrared Financial Messaging Point and Pay Profile (IrFM), ver. 1.0, Dec. 2003.

[7] Bluetooth Core Specification, ver. 1.2+EDR, Bluetooth SIG, Nov. 2003.

[8] Weiping Z.H.U, Dong WANG and Huanye SHENG, "Mobile RFID Technology for Improving M-Commerce". Proceedings of the IEEE International Conference on e-Business Engineering (ICEBE'05), March 2005, Shanghai, China.

[9] Raj BRIDGELALL, "Enabling Mobile Commerce Through Pervasive Communications with Ubiquitous RF Tags", Wireless Communications and Networking, WCNC 2003, Volume 3, Page(s):2041 – 2046, 16-20 March 2003.

[10] William Stallings, "Data and Computer Communications, Seventh Edition", Prentice-Hall, 2004.

[11] C.D. KNUTSON and J.M. BROWN, "IrDA Principles and Protocols:" The IrDA Library, Vol.1, MCL Press, 2004.

[12] H.R. DAMON, R.J. BROWN, and L. FAULKNER, White Paper, "Creating an End-To-End Digital Payment System," *IrDA Press*, Oct. 1999.

[13] Yingjiu Li and Xinwen Zhang, "A Security-Enhanced One-Time Payment Scheme for Credit Card", Proceedings of the 14th International Workshop on Research Issues on Data Engineering: Web Services for E-Commerce and E-Government Applications, RIDE-WS-ECEG'2004, Boston, USA , March 28-29, 2004.

[14] PayPal "Online Payment Processing", https://www.paypal.com/cgi-bin/webscr?cmd=_wp-pro-overview-outside.

[15] Sami IREN, Paul D. AMER and Phillip T. CONRAD, "The Transport Layer: Tutorial and Survey", *ACM Computing Surveys*, Vol. 31, No. 4, December 1999.

[16] Hasan Amca And Raygan Kansoy, "A Mobile Telephone Based, Secure Micro-Payment Technology Using The Existing ICT Infrastructure", Chinacom 2007: International Conference On Communications and Networking In China, 22-24 Aug. 2007, Shanghai, China.

Multiplexing Overlays on Bluetooth

Abdelshakour Abuzneid , Sarosh Patel
Viqar U.Mohammed , Varun Kumar Godula
Department of Computer Science & Engineering
University of Bridgeport
{abuzneid, saroshp, vmohamme, vgodula}@bridgeport.edu

Abstract- **This work aims at providing a proof of concept for multiplexing in Bluetooth by using traditional Time Division and Frequency Division multiplexing overlays upon frequency hopping spread spectrum (Bluetooth Modulation) by means of a functional simulation. This overlaid multiplexing technique can simplify the point to multi-point connections, especially when there are multiple Bluetooth devices in the vicinity and using the limited ISM spectrum. The available narrow bandwidth can be better utilized in point to multipoint connections using the proposed method.**

I.INTRODUCTION

Named after the Scandinavian King Harald Bluetooth, famous for uniting the countries Denmark and Norway during the 10th Century, Bluetooth was developed as a cheap short range communication protocol. This project Bluetooth was support by conglomeration of companies: Ericsson, IBM, Intel, Nokia, and Toshiba and rigorously promoted by many other wireless communication and portable devices companies. [1, 2]

Bluetooth uses the unlicensed Industrial, Scientific, and Medical (ISM) band around the 2.4 GHz (+/- 50MHz) frequency range. A variety of other home appliances and public devices operate in this unlicensed frequency band, such as, cordless phones, microwave ovens, garage door openers, broadcasting devices, etc. This causes a lot of interference and the disturbance in the small band. There are also other wireless communication technologies using the same Industrial, Medical & Scientific (ISM) band. The two protocols that operate in this band are the 802.11 also known as WiFi and Home RF.

Though initially Bluetooth was caught in a conundrum – there were very few companies to accept the technology as Bluetooth devices need other Bluetooth enabled devices to connect to, which very few. Bluetooth has really become popular with cost of Bluetooth chips falling due to great advancements in semiconductor and VLSI fields.

II. CURRENT BLUETOOTH STANDARD

The Bluetooth Version 1 standard gives specifications for voice and data communication over a radio channel with a maximum capacity of 1Mbps, though Version 2 enhanced Bluetooth claims to have attained an impressive data rate of 2.1 Mbps.

Bluetooth devices use the complete ISM band, while never transmitting from a fixed frequency for more than a very short time. This ensures that Bluetooth conforms to the ISM restrictions on transmission quantity per frequency. The available frequency spectrum is divided into 79 channels 1 MHz apart.

Bluetooth transmits at a low power (1mW) and is therefore designed for short-range use of less than 10 meters. Bluetooth can support an asymmetric link with up to 723.2 kb/s in one direction and 57.6 kb/s in the return direction, or a symmetric link with 433.9 kb/s in both directions at once. Bluetooth devices switch frequencies 1600 times per seconds (frequency hopping) making it almost immune to security risks and external interference even though the devices share the ISM band with many other wireless technologies.

Bluetooth devices are capable of both point-to-point and point-to-multipoint communications. Due to power limitations Bluetooth operates in three power classes depending on distance separation between the communicating devices.

There have been many versions of Bluetooth, the latest being Version 2.0. There have been significant improvements with every new version. Bluetooth 2.0

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 375–383.

has been designed to complement existing Bluetooth devices (backward compatible) and offers data transmission rates up to 2.1 Mbps [2].

Although Bluetooth is getting increasing popular one of the major limitations is a small bandwidth of 1 MHz. And this limited bandwidth has to be shared in case simultaneous of point-to-multipoint connections involving data transfer. This work proposes techniques to enable simultaneous transmissions from Master to multiple slaves in case of point to multipoint connections in a Piconet by effectively utilizing the available narrow bandwidth.

Before we proceed to the proposed method a clear understanding of Frequency Hopping Spread Spectrum (FHSS) and the Bluetooth architecture is necessary.

A shared Hopping code (a pseudo-random seed number) keeps the transmitter and the receiver tied, hoping to the same frequency channels simultaneously. [5]

Although the data rate is only a 1Mbps, a much larger bandwidth of 79MHz is needed. Transmission time is divided into 375 micro-second slots, with a new hop frequency being used for each transmission slot. Even although Bluetooth transmissions will occasionally collide with those from another device, this can be tolerated or recovered from with appropriate coding schemes. [3]

The frequency hopping spread spectrum makes Bluetooth inherently advantageous over other wireless techniques in terms of security and channel interference.

Figure 1: Bluetooth Frequency Hop Spectrogram

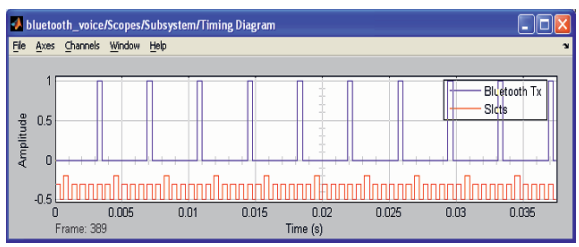


Figure 2: Bluetooth transmission bursts

III. FREQUENCY HOPPING SPREAD SPECTRUM (FHSS)

Frequency hopping is employed to avoid interfering with other devices transmitting in the already crowded ISM band. The frequency spectrum is divided up into 79 channels spaced 1 MHz apart starting at the low end of the 2.4 GHz ISM band. The transmitter switches channels 1 MHz wide 1600 times per second. Figure 1 shows the frequency hopping spectrogram.

Frequency hopping makes eavesdropping impossible, as the next hop is extremely difficult to guess. Even if the hop frequency channels can be known their dynamically changing hopping sequence is very hard to predict unless the Hopping code is known. This makes Bluetooth ideally suited in military applications, as Bluetooth transmission bursts change channels and also the dwell time is very short

in duration (625 ms), therefore hard to detect and even harder to jam.

Any channel interference will be short lived as the transmitter shifts to another frequency band in the next transmission slot and any packet lost can be re-transmitted [4, 5]. Please refer to Figure 3.

It also offers good resistance to multi-path path fading because the direct signal always arrives at the receiver first. Reflected signals follow a longer path and arrive later. By then the receiver may have changed frequency and no longer accepts signals on the previous frequency, thus eliminating interference between the direct and reflected signals. [3]

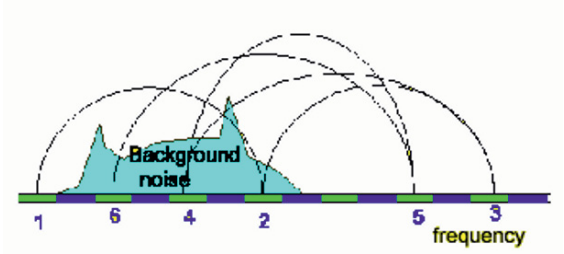


Figure 3: Hoping with background noise [4]

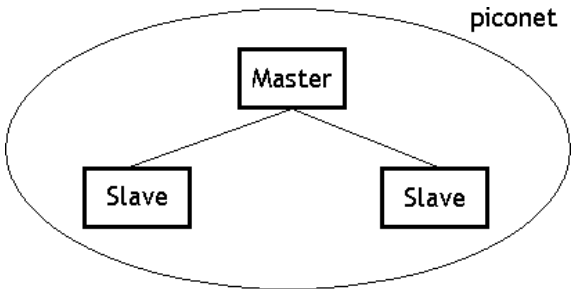


Figure 4: A Bluetooth Piconet

IV. BLUETOOTH CONNECTIONS

Whenever two or more Bluetooth devices are paired together one of them assumes the role of a Master and the rest are referred to as Slaves. A network of ad-hoc connections between a Master Bluetooth device and one or more Slave is called a Piconet. The current Bluetooth standard allows for up to seven active slaves to be paired with a Master device and up to 255 devices can be parked (connected but inactive) with the Master at any instant [6, 7, 8].

Devices alternately assume the role of Master in a Piconet in a cyclic manner. At any given instant of time, data transfer occurs between a Master and a Slave device.

Sometimes two Piconets can come together to form a Scatternet, where in a Master from one of the Piconets becomes the Master of the Scatternet and act like a bridge between the two Piconets [7].

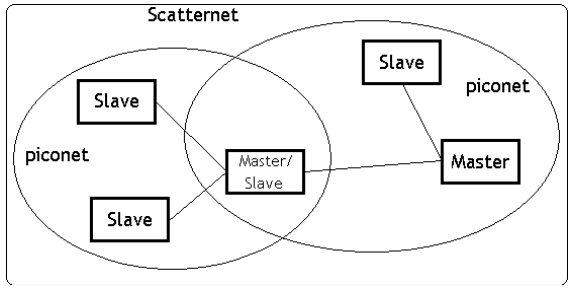


Figure 5: Bluetooth Scatternet

V. SIMULINK MODEL FOR OVERLAYS ON BLUETOOTH

As discussed above even though Bluetooth supports point-to-multipoint connects simultaneous data transfer with one then one device is seldom used. This is mainly due to the small bandwidth of 1 MHz available to the device. However, using the two suggested methods devices can seamlessly connect and have simultaneous data transmissions with more than one device.

When designing Bluetooth systems (as with any semiconductor devices), it is crucial to simulate and test them in the presence of interference from these other devices. System-level design tools like Simulink give engineers the capability to simulate the behavior of their devices and carryout such tests before commencing costly hardware and embedded software design. This allows the discovery of design flaws early in the development process while they are inexpensive to correct.

The suggested approach is a software overlay on the existing Bluetooth protocol and necessitates no change in the device hardware. The Simulink[1] Model for the Bluetooth protocol has been designed using the MATLAB communications blockset and signal processing tool box. The suggested overlay is modular in nature and is easily modifiable.

Simulink has been used to simulate the different scenarios. All the simulations consider a sample Piconet where a Master is connected with two Slaves. The intent of the simulation enables simultaneous data transfer between the Master and the two Slaves. Master acts as the transmitter the Slaves act as receivers. Both Master and Slave are the same (having a receiver and transmitter modules) but only differ by the roles they play within the Piconet.

A few simplifying assumptions have been made to enable faster simulation. Every Bluetooth has the same functional capabilities, consisting of the transmitter and receiver modules. The simulation takes into consideration atmospheric effects such as path loss in the medium and interference is implemented using Gaussian white noise model.

The above Figure 5 shows a Piconet consisting of a Master device and two Slave devices. The simulation aims to transmit data simultaneously to both the slave devices using Bluetooth. The Master device acts as the transmitter and the Slave acts as the receiver, neglecting the acknowledgements it transmits.

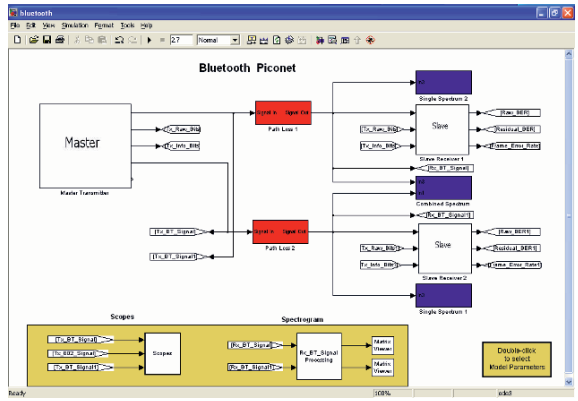


Figure 6: Piconet model in Simulink

The transmitted data packets can be viewed using the spectrogram graph and also the device timing diagram. The Master shares a different set of Hopping codes with each of the two Slave devices, therefore each of the two Master-Slave connections are independent and the files transferred are also different.

The generation of the hopping code and the process of establishing and maintaining a Bluetooth connection is out of the scope of the work presented in this paper. The hopping code for an overlaid Bluetooth connection will be different from that of a regular Bluetooth connection, since additional parameters need to be conveyed in case of the overlaid connection.

VI. TDM OVERLAY ON BLUETOOTH

This overlay module implement the Time Division Multiplexing (TDM) module on the Bluetooth connections. Bluetooth protocols have six slots to transmit data. The transmitter and receiver are synchronized by means of the shared hopping code.

In this case the Master device is synchronized with two devices with the same hopping code. But Master transmits data to each of the Slave at different slots. This is because using only one transmitter module the Master can not jump to two frequencies at the same time. Hence, while it is at the one frequency it transmits data to both the devices but at different time slots. The dwell time at hopping frequency can be accordingly adjusted so as to provide for sufficient guard time between the two data transmission. More importantly the two receivers need to be properly synchronized with the transmitter. The encoding and modulation of data is the same as that used with the Bluetooth standard protocol.

From the receiver's point of view it is no different from a simple Bluetooth connection. The receiver has to hop to the frequencies as dictated by the hopping code. Master needs to communicate this to the receiver, as to which time slot it should expect to receive the data.

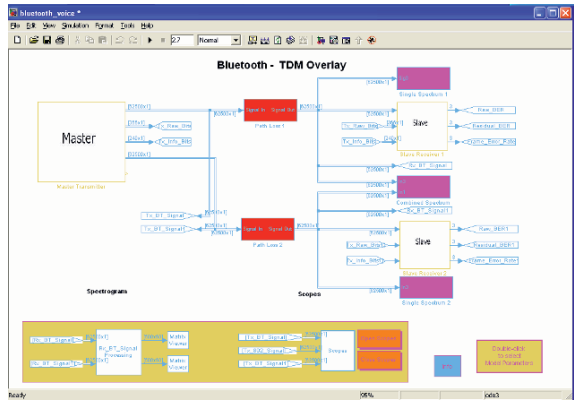


Figure 7: TDM Overlaid Simulink layout

As seen in Figure 6 and 7 the hopping sequence for both the channels, channel 1 (Master – Slave 1) and channel 2 (Master – Slave 2) is the same. But, the data is transmitted at different time slots. This can be clearly seen in the timing diagrams. The Master device transmits data to Slave 1 at every 4th time slot (Figure 8) and it transmits data to Slave 2 at every 1st time slot (Figure 9), within each hop. The two transmissions have a time difference of 3 slots between them. Therefore, there can be 6 simultaneous connections with the master each using a time slot, provided the dwell time at each frequency is long enough and there is appropriate guard time between transmissions for the receivers to know the difference.

Figures 8 and 9 again confirm the fact that both the connections shared the same hopping sequence. The figure show a Bluetooth transmission burst from the Master device in the frequency domain.

Figure 8: Hopping Spectrogram of Connection 1 (X-Axis-Time, Y-Axis-Frequency)

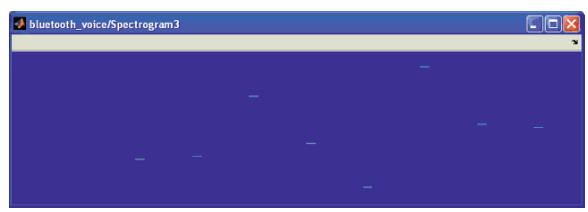


Figure 9: Hopping Spectrogram of Connection 2 (X-Axis-Time, Y-Axis-Frequency)

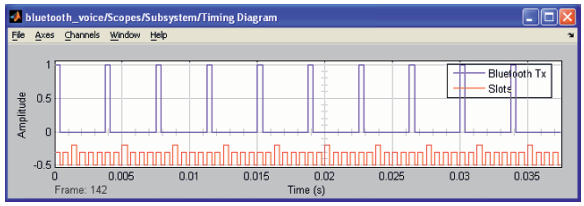


Figure 10: Timing Diagram of Connection 1

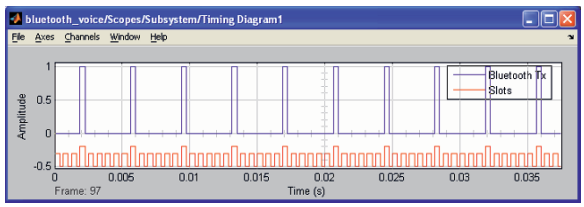


Figure 11: Timing Diagram of Connection 2

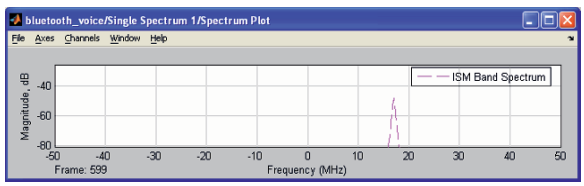


Figure 12: Connection 1 transmission burst

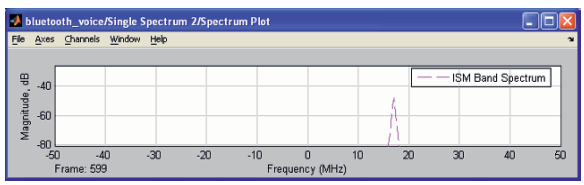


Figure 13: Connection 2 transmission burst

In special cases and additional facility can be incorporated in the TDM which will shared the hops between the two links in proportion to the size of the data transfer needed. For example, out of the 79 hops per second say 30 hops are dedicated to the connection with Slave 1 and the rest 49 are dedicated to the connection with Slave 2. This has to be negotiated while setting up the connection and the hopping sequence has to be then adequately generated. This new hopping code will tell the receiver which slots to listen to and which not.

VII. FDM Overlay on Bluetooth

In the case of and FDM overlay, let us assume that the ISM band is equally split into two equal bands ISM1 (lower half) and ISM2 (upper half). The Master can communicate with the two Slave devices independently with same hopping code by using these two half bands.

The hopping sequence for both the connections is the same. The only difference is that they operate in different halves of the same band, hence do not interfere.

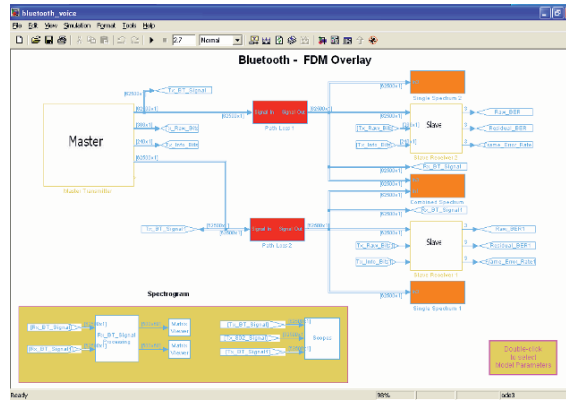


Figure 14: FDM Overlaid Simulink layout

Having the same hopping sequence eliminates the need for additional resources for generating second hopping frequencies. Assuming the hop generator generates frequencies for the lower band, the upper band frequency will be a constant higher than the lower band frequency. Hence, in this case the dwell time in each frequency is reduced to half the original. Consequently the in terms of hardware the frequency oscillator will have to switch frequencies twice as fast. Also, the hopping sequence's random seed should be chosen such that the consecutive frequencies generated are not to close, as the available total available bandwidth is half of the ISM band. Though the bandwidth at any available hop is still the same at 1 MHz.

Depending on the necessity of the connections the dwell time can be altered to there by reducing the number of hops per second. Otherwise the number of hops per second can be reduced if the dwell time needs to be kept constant.

Figure 15: Connection 1 FHSS (X-Axis-Time, Y-Axis-Frequency)

Figure 16: Connection 2 FHSS (X-Axis-Time, Y-Axis-Frequency)

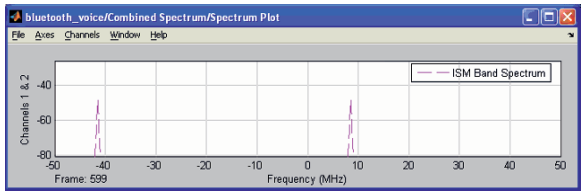


Figure 17: Data transmission bursts in the ISM Band

Figure 16 shows two Bluetooth transmission bursts in 100 MHz ISM band separated by 50MHz. This overlay helps the Master device maintain two data connections simultaneously with two slaves by splitting the available band into two. Both the files transmitted to the Slaves were received successfully at the end of the simulation.

For specialized applications where one connection dictates faster data transfer rate than the other, the ISM band can be unequally divided. Allocating the larger band to the connection demanding the higher data transfer rate, this in turn supports more hops per second.

VIII. CONCLUSIONS

In this paper we have presented a novel approach to implement TDM and FDM overlays on the Bluetooth protocol. These overlays help the Master maintain more than one data connection with the Slaves simultaneously.

IX. FUTURE WORK

Bluetooth technology is a very promising field of research and development, especially considering the fact that it is an evolving technology currently in its nascent stages. In the near future we will be concentrating our efforts mainly on the following tasks:

1. Developing algorithm for generating the hopping code for an overlaid connection.
2. Protocol for establishing an overlaid connection between the Master and Slave device.
3. Implement a working model in hardware.

REFERENCES

[1] The Bluetooth Special Interest Group (SIG) (https://www.bluetooth.org/apps/content/)

[2] The Bluetooth official website (http://www.bluetooth.com)

[3] “Computer Networks”, Andrew S. Tanenbaum, Fourth Edition, Peterson Education.

[4] “An Performance with ISM-Band Narrowband Interference” by Dr. Earl McCune.

[5] “Bluetooth demodulation algorithms and their performance”, by Roel Schiphorst, Fokke Hoeksema and Kees Slump.

[6] “Performance of Bluetooth Bridges in Scatternets with Limited Service Scheduling”, Vojislav B. Misic and Jelena Misic, Mobile Networks and Applications, Volume 9, Issue 1 (Feburary 2004), Pages: 73 - 87.

[7] “Bluetooth Scatternet Models”, McDermott-Wells, IEEE Potentials Magazine, December 2004 - January 2005.

[8] “Performance aspects of Bluetooth scatternet formation”, Miklos, G.; Racz, A.; Turanyi, Z.; Valko, A.; Johansson, P., Mobile and Ad Hoc Networking and Computing, 2000. MobiHOC. 2000 First Annual Workshop on Volume 9, Issue 1, 2000 Page(s):147 – 148.

The Problem of Predicting the Data Transmitting Delay in the Network with the Self-Similar Nature of Traffic, for the Purpose of Improving the Real-Time Conferencing

I. Sychev
Hasso-Plattner-Institut, Room C 1.10
2 Prof.-Dr.-Helmert Str.
Potsdam, 14482 Germany
(igor.sychev@hpi.uni-potsdam.de, sychov@tfh-berlin.de)

Prof. Dr. A. Polze
Hasso-Plattner-Institut, Room C 1.18
2 Prof.-Dr.-Helmert Str.
Potsdam, 14482 Germany

D. D. Gazzaev
Amur State University
21 Ignatevskoe shosse
Blagoveschensk, Amur Region, 675027 Russia

Prof. Dr. C. Kordecki
Labor für Prozess- und Systemtechnik FB VI
10 Luxemburger Str.
Berlin, 13353 Germany

I. A. Sycheva
Amur State University
21 Ignatevskoe shosse
Blagoveschensk, Amur Region, 675027 Russia

***Abstract*- This publication gives new results in applying the theoretical knowledge based on the Laplace-Stiltes transform. The main purpose is to predict packets transmitting delay, in a network based on the Internet technology. A new method for modeling the real-time networking process is designed.**

I. Introduction

The main ideas of the researched problem were presented at the Fraunhofer Institute Computer Architecture and Software Technology (Fraunhofer FIRST), Berlin, Germany and at the University of Applied Sciences Berlin, in the internal interviews.

Researching the Real-Time Network Applications, scientists defined the problem: how to predict the possibility of the Real-Time conference based on the Internet technology network? Even if the bandwidth of the connection is good, the packets transfer delays might prevent real-time conferences.

Nowadays, multimedia systems for the real-time conferencing become more and more challenging – ranging from the High-resolution conferences for the big halls, to the digital planetarium installations, and it is quite possibly that, in the nearest future, to the digital real-time planetarium which shows the video from the digital telescope which runs over the orbit around the Earth (or video from distributed space observatories). The other obvious alternative of the video source for these systems might be programs for 3 dimensional graphics (3D graphics). The example of a big installation for a video output is presented on Fig. 1.

Fig. 1. The video installation

But, in most cases, for above-mentioned systems we need big and expensive computer clusters (to generate or compress the video flow). We want to use the video sources and the video outputs remotely with the best possible quality and we want to know if a real-time conference is possible with the Internet technology connection that we have. We want to know the most important set of the connection parameters and we want to get the simulation of presented problem.

Below, part II presents the problem-relative solutions. Part II gives information about the current solutions in the similar area to the presented research. Common information about prerequisites for research is presented in part III. This part presents the most important references, which are the base of current research. Part IV describes the main idea of the algorithm for predicting the packets transfer delay. The algorithm is based on the Laplace transform in

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 384–388.

the shape of Stiltes integral. Part IV presents the new effective algorithm formulated from wide mathematical theory about the Laplace transform properties. Part V is the key part as it presents the main results of the simulation. Part V shows to us opportunities for practical application of theoretical knowledge described in part IV. And, therefore, part V is the major contribution of this paper. Part VI provides the conclusions.

II. The Problem-Relative Solutions

At the present moment lot of products are produced under the title: "real-time conferencing". In this paper, real-time conferencing is the process with the defined properties:

- the conference is produced between two remote computer systems (between two participants);
- the systems produce data blocks;
- a time required for generating or obtaining the data block is available for measurement;
- a data block must be transmitted by the Internet-technology network;
- the amount of frames or packets which has the fixed length and required for transferring the data block (using the Internet technology) is available for counting;
- the amount of data blocks is available for counting;
- the timestamps of the packets arriving (to the receiver system) can be measured;
- the time required for processing the obtained data and the time required for the output device is also available for measurement.

The process with the abovementioned properties can be realized by the protocols:

- a transport protocol for real-time applications (RTP) [1];
- a real-time streaming protocol (RTSP) [2];
- a RTP control protocol (RTCP)[1];
- a Zimmermann RTP(ZRTP) [3];
- a resource reservation protocol (RSVP) [4];

and by the "specification of guaranteed quality of service" (QoS) [5].*

But, presented research has one more purpose: to produce as much control as possible (if and only if it is possible) over the network components for the purpose of obtaining the best possible quality of the real-time process.

It is not possible to use most of presented protocols for above-mentioned purpose. RTP, RTCP** and ZRTP gathers statistics on a media connection and information such as bytes sent, packets sent, lost packets, jitter, feedback and round trip delay. An application may use this information to increase the quality of service perhaps by limiting flow, or maybe using a low compression codec instead of a high compression codec [1]. So, the application should solve the frequency predicting problems and operate with the data and channel. RTSP takes advantage of streaming which breaks data into many packets sized according to the bandwith available between client and server [2]. But, it does not control the frequency of packets. The specification of guaranteed quality of service [5] says that the fixed delay is a property of the chosen path, which is determined not by guaranteed service but by the setup mechanism. Only queuing delay is determined by guaranteed service.

In this case, only RSVP is able to provide the required mechanisms. Presented research is focused on possible improving the functionality of RSVP for the rate-sensitive traffic. This paper is focused on the mathematical model for the real-time conferencing process, and on the predicting of the possibility of real-time conference with the fixed quality.

*It is possible to find more protocols that fit to the definition of conferencing process. This paper provides only the example for most famous protocols.

**RTP and RTCP exist as an Internet Standard.

III. Prerequisites For Research

A lot of recent scientific papers about the network traffic describes the self-similar nature of traffic. Basic ideas and results are presented in [6]. The research presented in [6] is verified by many authors, including [7-9].

This article presents the new mathematical model of the data blocks transmitting delay in the network with the self-similar nature of traffic.

For obtaining the expected value of the data block delay we use mathematical apparatus presented in [10]. The author of [10] applies the theoretical knowledge presented in chapter II of [10] to the distributed data base management systems (DBMS) algorithms, while our paper is focused on the networking and shows the new application of the theoretical knowledge.

In the presented article we use the definition of the process of data transmitting as the process described in details in [11].

Let's formulate an algorithm, using [10] in chapter IV of this paper.

IV. Algorithm For Determination Of The Time Of Receiving-transmitting A Data Block

To determine and predict the accurate time of receiving-transmitting a data block let's formulate the algorithm "A" from [10]

A1. Find the cumulative distribution function (CDF) *F(t)*; where *F(t)* is the CDF of random time of one packet transmitting.

A2. Find the Laplace transform in the shape of Stiltes integral (further Laplace-Stiltes transform) from *F(t)*.

$$\Psi(\omega) = \int_0^\infty e^{-\omega t} dF(t) \tag{1}$$

A3. Experimentally determine the probabilities P_i that we need n packets for transmitting of the data block.

A4. Find the generating function of the packets count*.

$$\gamma(z) = \sum_{i=0}^{\infty} P_i \cdot z^i \tag{2}$$

A5. Insert (1) into (2) to get the Laplas-Stiltes transform of the time to transfer all packets.

*In the strict sense the infinity limit in the sum at (2) makes (modifies) from algorithm a calculation method (in computer science an algorithm without finite steps property called - calculation method [16]).

$$T(\omega) = \gamma(\Psi(\omega)) \tag{3}$$

A6. Find the first derivative from (3).

A7. To use the property of Laplas-Stiltes transform [10]

$$T^{(n)}(0) = (-1)^n \cdot M(\xi^n) \tag{4}$$

where $T^{(n)}(0)$ – is the n derivatives $T(S)$, when $S=0$, $M(\xi^n)$ – the n^{th} initial moment of the random value ξ; $M(\xi^n)$ – expected value.

V. The Matematical Model Based On The Algorithm A

Obviously, the simulation based on algorithm A depends on the $F(t)$. We propose to use well-known approximations of network traffic according to the chapter III of this paper. The approximation ideas and models are presented in [12].

Undetermined time fractals have been widely used to model a number of phenomena in diverse fields from biology to finance and networking*.

In chapter 9 of [13] the set of undetermined time fractals, including the fractional Brownian motion (FBM), is presented. The existence of FBM was established by Benoit B. Mandelbrot and J. W. Van Ness [14].

FBM takes into consideration the Herst parameter, nowadays the Herst parameter is the main Self-Similarity metric. Description of Herst parameter mathematics runs out of this paper frames, but it is sufficiently represented in [6-9].

The base of FBM is the Gaussian Random Walk (GRW) [13]. Let's see how algorithm A functions with GRW apparatus.

The first derivative from CDF is the Probability density function (PDF). It is easier to use (1) transformed to the shape (5), when here and down the paper $F(t)$ is the PDF.

In this case, (1) transforms to

$$L(S) = \int_0^{\infty} e^{-St} F(t)\, dt \tag{5}$$

and we get the Laplace transform.

The PDF of GRW is the normal distribution**

$$F(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma} \cdot e^{\left(-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2\right)} \tag{6}$$

where μ is the expected value, σ is the variance and x is the variable of PDF.

*Not all undetermined time fractals fit perfectly to build the simulation using the algorithm A; nowadays we did not find solution for undetermined Wierstrass-Mandelbrot time fractal [15].

**The normal distribution, also called the Gaussian distribution, is an important family of continuous probability distribution, applicable in many fields.

Let's $\mu = 0$, according to [13].

Using (5) we get the Laplace transform of (6)

$$L(S) = \frac{\sqrt{\pi}\cdot\sigma\cdot e^{\left(\frac{\sigma^2\cdot S^2}{2}\right)}\cdot\left(1 - erf\left(\frac{\sqrt{2}\cdot\sigma\cdot S}{2}\right)\right)}{\sqrt{2\pi}\cdot\sigma} \tag{7}$$

Then, we use (3) and (4) to get the expected time of transmitting the data block (First, we obtain the sum similar to (2) inserting $L(S)$ from (7) instead of z in (2). Then we get the first derivative from this sum, and afterward we assign $S=0$; so we get $T^{(1)}(0)$. Finally we multiply the sum to $(-1)^1$. So, we get (8)).

$$M(\xi^1) = (-1)^1 \cdot \sum_{i=0}^{\infty} i \cdot P_i \cdot \left(\frac{1}{2}\right)^{i-1} \cdot \left(-\frac{\sigma}{\sqrt{2\cdot\pi}}\right) \tag{8}$$

P_i – is the probability that for one data block we need i packets to transmit the data block.

$M(\xi^1)$ – is the expected time for transmitting one data block. σ – is the parameter of the packets transfer time variance. The longer time period for measuring σ, the more accurate is σ. We propose to calculate σ in the first seconds of data transferring and then to renew the σ value in the user-defined frequency.

If I is the quantity of probabilities P_i, then in the experiment with networking equipment, I is always a natural finite number. In this case, P_i when $i>I$ is 0 and all next members of sum (8) become 0. To transform the algorithm A to an algorithm in the strict sense, we give the finite steps property for A, changing the infinity symbol in the sum (8) to the finite natural number I.

The data measured in the real-time conference experiment is presented in table 1. The result of (8) with different σ is presented on Fig. 2. Fig. 2 shows that the GRW (non-fractal) network process depends only on σ and probabilities P_i. The model (8) provides the rule: the higher the process variance, the longer is the time of the data block transferring. Taking into account the results of [11] we can make an assumption about the future avoiding the self-similar models in preference to the random process models.

The model (8) does not take into account the possible self-similar nature of traffic. For this purpose we get (9) the PDF of FBM from [13].

TABLE I
The Data Fragment

Probability Number	Probability value	Packets count
1	0,011	1
2	0,012	2
3	0,303	3
4	0,510	4
5	0,163	5
6	0,001	6

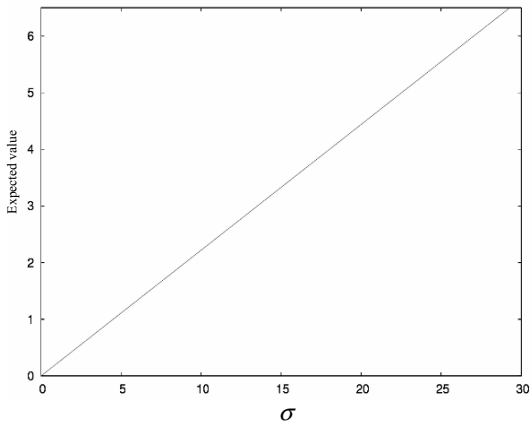


Fig. 2. The dependence of the data block transmission expected value to the process variance

Let's, $P(A/B)$ – is the conditional probability of event A, when the event B is established. Lets, $X(t_k)$ – is a random value in the moment of the time t_k, k – is a natural number. We are interested in a process such as $P(\mathrm{X}(t_k) \le x_k / \mathrm{X}(t_{k-1}) = x_{k-1})$, where: $t_1<t_2<...<t_{k-1}<t_k$; here and down the paper, t_1 – is the initial time of the process, t_2 – is the time moment of obtaining the probability of the event A (this notation is equal to [13]). Lets $T_\Delta = t_2 - t_1$.

σ – is the packets transfer time variance, x is the variable of PDF, H – is the Herst parameter.

$$F(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma\cdot {T_\Delta}^H}\cdot e^{\left(-\frac{x^2}{2\cdot\sigma^2\cdot {T_\Delta}^{2H}}\right)} \qquad (9)$$

If $\sigma > 0$ and $0<H<1$, then the Laplace transform of (9) is

$$L(S) = \frac{e^{\left(\frac{\sigma^2\cdot S^2\cdot {T_\Delta}^{2H}}{2}\right)}\cdot\left[1-erf\left(\frac{\sqrt{2}\cdot\sigma\cdot S\cdot\sqrt{{T_\Delta}^{2H}}}{2}\right)\right]}{2} \qquad (10)$$

Using (3) and (4) we get the expected time of transmitting the data block for FBM.

$$M(\xi^1) = \sum_{i=0}^{I} i\cdot P_i\cdot\left(\frac{1}{2}\right)^{i-1}\cdot\left(\frac{\sigma\cdot {T_\Delta}^H}{\sqrt{2\cdot\pi}}\right) \qquad (11)$$

$M(\xi^1)$ – is the expected value of the data block transmission time. Fig. 3-5 presents the results of mathematical modeling.

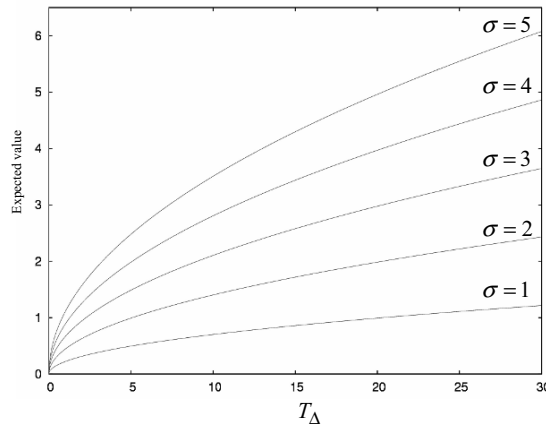


Fig. 3. The results of calculating the expected value of the data block transmission time, when H=0.3

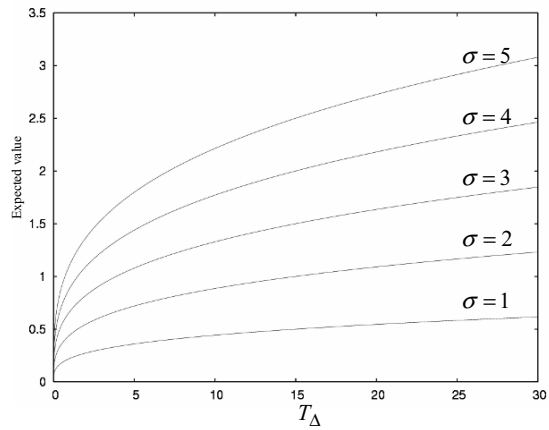


Fig. 4. The results of calculating the expected value of the data block transmission time, when H=0.5

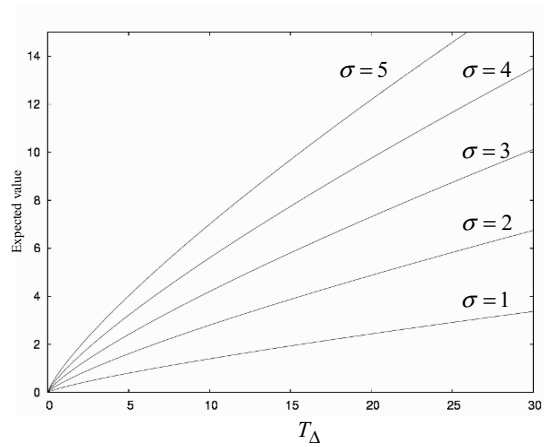


Fig. 5. The results of calculating the expected value of the data block transmission time, when H=0.8

Fig. 3-5 shows the results of the simulation, where the time of the data block transfer delay is obtained. Obviously, Fig. 3-5 shows that the dependence of the process on the previous events makes the forecast more problematic, and, in this case, the expected value of the block transmission time is increasing. If the variance σ is low, then it is possible to make the forecast to the longer T_Δ. Fig 2-5 shows the process with the variance up to 5; that is not typical for the modern networks. Fig 2-5 shows the process that is similar to a connection with a really bad quality, for the purpose of demonstrating the mathematical apparatus on the small values.

The software based on presented research might function in the following way: T_Δ is defined by the system user; P_i, I, σ are measured during the short-time test or during the conference. Then the software calculates the expected value. If the expected value is too high, than the system proposes to decrease the variance σ or to setup the lower T_Δ. In modern networks it is possible to get the different σ switching between the available routes in the routing map or upgrading the existed channel.

Presented model does not depend on the transport layer of the networking algorithm. If the packets loss exists and is controlled by the control protocol (or by the real-time protocol, like RTSP), then the variance σ grows. If the packets loss is not controlled, then it means that the data fragments loss is allowed by the system user.

VI. Conclusions.

Obtained knowledge can be effectively used for enhancing the quality of service of the network based on the Internet technology. The main application of the presented model is the real-time conferencing and a video flow transferring.

Conducted work gave the following outcomes:

– new method of the networking traffic modeling is designed;

– the new method takes into account all the basic properties of the modern Internet-technology network;

– this paper shows the importance in selecting the model;

– the opportunity to compare obtained data from the different models with the identical input values.

Presented research gives the opportunity to implement obtained knowledge to the software. The software tests and implementations is the main subject of the future research.

References

[1] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, “RTP: A Transport Protocol for Real-Time Applications”, RFC 3550, July 2003.
[2] H. Schulzrinne, A. Rao, R. Lanphier, “Real Time Streaming Protocol (RTSP)”, RFC 2326, July 2003.
[3] P. Zimmermann, A. Johnston, Ed., J. Callas, “ZRTP: Media Path Key Agreement for Secure RTP”, Internet-Draft, July 9, 2007.
[4] R. Braden, Ed., L. Zhang, S. Berson, S. Herzog, S. Jamin, “Resource ReSerVation Protocol”, RFC2205, September 1997.
[5] S. Shenker, C. Partridge, R. Guerin, “Specification of Guaranteed Quality of Service”, RFC2212, September 1997.
[6] Will E. Leland, Murad S. Taqqu, Walter Willinger, Daniel V. Wilson, “On the Self-Similar Nature of Ethernet Traffic”, 1993.
[7] Vern Paxson, Experiences With Internet Traffic Measurement and Analysis, ICSI Center for Internet Research International Computer Science Institute and Lawrence Berkeley National Laboratory, 2004.
[8] Peter Haga, Peter Pollner, Gabor Simon, Istvan Csabai, Gabor Vattay, Self-generated Self-similar Traffic, Communication Networks Laboratory, Eotvos Lorand University, 2004.
[9] Sergejs Ilnickis, “M/M/1 and G/M/1 systems with a self-similar input traffic”, 2004.
[10] A.D.Plutenko Execution time assessment of subqueries to relational databases, Computational Technologies vol. 5 no 4 Novosibirsk, 2000 [Reviewed by “Zentralblatt für Mathematik” and by “Mathematical Reviews”].
[11] I.V.Sychev, The Problem of Accurate Time Measurement in Researching Self-Similar Nature of Network Traffic, IEEE Conference Publishing, 2006.
[12] T. Mikosch, S. Resnick, H. Rootzen, A.Stegeman, "Is network traffic approximated by stable Levy motion or fractional Brownian motion?", the Annals of Applied Probability Vol. 12 No. 1, p. 23–68, 2002.
[13] R. M. Crownover, Introduction to Fractal and Chaos, London, 1995.
[14] Benoit B. Mandelbrot and J. W. Van Ness, Fractional Brownian Motions, Fractional Noises and Applications, SIAM Review, Vol. 10, No 4, 1968.
[15] I. E. Shishkin, Modeling and analysis of the fractal objects in the space and time, Ekaterinburg, 2004.
[16] Donald E. Knuth, The Art of Computer Programming vol. 1, 1997.

Guidelines for Constructing Robust Discrete-Time Computer Network Simulations

John Richter
Rhodes University

Barry Irwin
Rhodes University

September 2007

Abstract- **Developing network simulations is a complex task that is often performed in research and testing. The components required to build a network simulator are common to many solutions. In order to expedite further simulation development, these components have been outlined and detailed in this paper. The process for generating and using these components is then detailed, and an example of a simulator that has been implemented using this system, is detailed.**

I. INTRODUCTION

Simulation is a common technique used in research. By building a simulacrum of some real-world concept, it is possible to generate accurate predictions for the outcome of an event. Simulation is particularly useful in scenarios where the object of the simulation is hard or impossible to control due to expense, rarity or distance. Because simulation lets a developer create a 'virtual' equivalent of the object, it can be designed entirely under the user's control, with user-specified parameters and environment. Once common use of simulation that uses this principle is the design of simulated networks of computers.
In this paper, design considerations for the construction of network simulators are shown. Through the use of the optimizations mentioned and by addressing the problems and challenges in the design phase of the simulator, development of such a system can be made more efficient.

A. Networks and Simulation

The Internet is an electronic, virtual network of interconnected, heterogeneous hosts. Because of its massive scale and diversity, it is often considered difficult or impossible to determine the success or efficacy that any interaction can have with it (such as the implementation of a protocol or application, the rapid spread of malware, or the sudden loss of a large, central portion of the network's connectivity), aggravated by it's constant change.

One method of improving our understanding of the Internet, and thereby allowing us to better gauge effects on the internet, is simulation. Network simulation allows the operation of a simulacra of the Internet (or some subset of it), which could then have a translated equivalent of an effect imposed upon it to influence the state of the simulated system. Using this simulacrum, one could then determine the degree to which the system is changed because of the effect, and thus have a measure of its efficacy.

While concerns have been expressed regarding the effectiveness of a full-Internet simulation [1], it is possible to address some of the concerns mentioned, resulting in a high-quality (but nevertheless imperfect) simulation that is still very useful for determining consequences of change.

B. Scope

The scope for this document is the construction of network simulations, with the additional requirement that the time-system for the simulation is limited to a discrete set of times.
A decision that needs to be made in the construction of a simulator is the choice between analogue time simulation and discrete time simulation. This paper will present a structure for creating simulations that focuses on discrete time simulation - analogue time simulation is out of the scope of this document.
Various comments have been made on simulation, most practically on the nature of the packet system of the simulator. In [2] the difference between the two forms of simulation is shown in the dichotomy between analytical and packet-level simulations. By simulating in ways that translate more naturally into programmable terms that are similar to the original, the simulacrum models the original more accurately at the expense of efficiency.

The focus of this work is for simulations capable of being executed on a single desktop computer. While grid computing and mainframe computers can provide much more precise and detailed results, they are not commonly available for research work, and so are not the focus of these guidelines (though this paper may still prove useful).

Finally, the simulation framework was developed for robust simulations, where the system is broad in order to enable the development of a diverse range of simulations. Simulations such as Internet topology and routing [4], scalability [5], applying theoretical mathematical models of the Internet to a simulated network [6], testing new sorts of simulations [7], should all be possible within the bounds of this framework.

C. Overview

In Section II, requirements for a network simulation will be detailed, breaking the development down into two perspectives: the views and needs of the user of the system, and the details required by the developer. In Section III, challenges to practical use and operation of the simulator are stated, and algorithmic solutions suggested. Section IV

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 389–394.

covers implications of this work. Possible extensions to the basic system are considered in Section V. Finally, in Section VI, a final overview and summary are made.

II. Simulation Design

Various approaches and viewpoints must be considered when constructing any complex software. In order to reduce the complexity of a simulator, we shall consider the requirements, first from the point of view of a user of the simulator (to understand from a high-level perspective), and then from a mechanical perspective (to understand the fundamental structures and internal working of the simulator).

A. User Components

Considering the system from the user's perspective, the simulator must have several vital components. Firstly, it must be able to take input describing the internet, and the state that it is in, with special cases noted (such as a host infected with malware, or some vital areas shut down). This implies that the network will require some sort of data structure to represent the state of the network.
Then, as simulated time passes, the state would need to change. This implies some component that will apply changes to the aforementioned network state structure.

As the user will require some sort of feedback for the system to be useful, a component that can review the state of the network structure is also necessary.
Finally, related to the review system, some sort of information aggregator might be called for. Because simulations of networks are often large, the amount of data such a simulation produces will be impossible to understand without some mechanism to collate and aggregate the information derived from the simulator.

A.1. Network State

The state of the network is a core structure in the simulation. At simulation creation-time, the network state is entered by the user, and is then altered as the simulation proceeds. As it changes, behaviour can be observed at intervals to determine the current state of the system. Finally, at the end of the simulation, the network state should be presented to the user for final inspection.
Throughout the entirety of the simulation, the state structure is the chief focus of the user. As it has been specified that the system is a network simulator, we can assume that the states will represent states of hosts upon the network. This can include details such as operating system of a host, whether it is infected with malware, or whether the host is powered on and active.

The state structure does not necessarily only specify details about hosts. It can also include information about other aspects of the network, such as the connections between hosts (with details such as bandwidth capacity and packet drop rate) and packets in outgoing or incoming queues (packet size, protocol, etc.).

A.2. State Entry

When the simulator is initialised, the user must input information regarding the state of the network that is to be simulated. This can take the form of a complete document that has been prepared, or command-line parameters.

As simulations exist to test and derive information from events or changes that will alter the status quo, it is possible to build a standard entry state for a simulator which can then be edited to take into account any changes the user wishes to impose on the system.
If the simulation is large, it is recommended that the input network state be generated programmatically. Simulating even a small subset of the entire Internet will require a great amount of data entry. Instead, using random generation and statistics to derive a detailed model of the Internet will take little time.

A.3. State Alteration

The purpose of a simulator is to alter the state of the input data in order to generate some result. The alteration stage is the most important action of the simulation, taking the network state that has been entered into the system during the entry phase and applying adjustments to represent how the system will change in a real system.
State alteration focuses on the changes in the network state, so specifics such as details of the hosts and network connectivity are changed, acting upon the network state component of the simulation. As the network state changes from that entered by the user, the simulation generates useful information

A.4. State Review

In order to extract information from the simulation system, some form of review of the network state is required. This should be able to be filtered, as a full output of every node in a large networked system could be incredibly large. It is suggested that certain details associated with nodes (such as malware infection) can be used as a filtering system, listing only hosts in the network that are interesting in terms of the simulation.
Output can be in a variety of formats, and it is often useful to create a program that is capable of parsing this information, partly to summarise and comment (detailed in the following section), and also to find abnormalities or specific details that summaries and statistics may overlook.

A.5. Result Commentary

Closely tied to the state review component, commentary and information aggregation is an important part of a simulation. Because a simulation can generate a large quantity of data (especially if the simulated network is large, or if the duration of the simulation is long), it is often difficult or impossible to manually extract useful information from the outputs, even if they have been filtered in the state review stage.
Other aspects of this component can include inference and statistical engines, deriving more useful information and concepts (especially if they can rely on non-filtered data from the state review component), and allowing the user a deeper insight into the execution of the simulation.

A.6. User-Focused Simulation Review

If we have known network state α, and the simulation is at time t, we have some alteration function for our network f(α, t), such that f(α, 0) = α, and f(α, 1) ≈ α , then, as $t \to \infty$, f should diverge from α. Assuming the simulation ends at time s, then we must be able to derive useful information from network state with associated alteration function f(α, s). Unfortunately, state α is often large and complicated, so some function, g(f(α, s)), needs to be run to parse the network state and extract useful information. To paraphrase, the system begins in a state known to the user (presumably as they entered it). As the system executes, it begins to diverge more and more from its initial state, and, upon termination of execution, needs to be able to present to the user useful information about the state, instead of a complete output.

B. Mechanical Components

Considering the simulator from a structural and mechanical point of view contrasts it to Section A. In this section, we consider the more practical components that will be required to be implemented, particularly as a network simulator.

A network (N) can be said to be comprised of a set of hosts, n, a set of connections between those hosts, c, and information running over the communication structure, p, moving from some hosts to other hosts. This can be expressed as N (n,c,p). Focusing in turn on each of those parts, it can be seen that a network simulation will need to have structures in place that can hold information about a series of hosts, and a structure (optionally a part of the host structure) that keeps a record of the connections between the hosts. It will also have to have some part that can simulate the transfer of information between the hosts (or nodes), over the connections. These will be termed packets of information, though they do not necessarily imply a packet-switched network.

As the function of a simulator is to take an initial state and alter it subject to various actions, a subsystem of the simulator must be in place that alters the state and communication components. The two components of this subsystem are the execution engine and the scheduler. These two tightly-coupled components operate upon the network states to change them with pre-defined methods.

B.1. State Components

The state components of the network represent routers, computers, and other devices with network interfaces. Devices with network cards are diverse, firstly between physical types (such as wireless network connections in an Aibo robotic dog, and a networked desktop PC), and within specific types (such as the differences between a PC running Windows, and one running Linux). Because this diversity will often affect the simulation in some way (such as malware spread not affecting the robotic dog), it is important when designing your components to know which parameters you need to include in the hosts.

Common parameters to consider are physical type of hardware, whether the device is operational, the operating system, and installed software and services.

This component corresponds with the network state component from Section A; however it is more focused on the nodes of the network.

Nodes on the network can be addressed formally using the numerical set-subscript notation, n_a. The first node on a network (usually the router) would be represented as n_1. For purposes of usability, the subscript can be used non-numerically to represent a filter. n_{pc}, for instance, could represent all hosts on the network that are Personal Computers, and $n_{infected}$ would represent all nodes that have been infected by malware.

B.2. Communication System

The nodes in the state components need to be able to send messages to other state components. Whether these messages are interpreted or are meaningful is dependant on the simulation, as some simulations will focus on concepts like available bandwidth, throughput or latency where the contents of the packets are irrelevant. This means that some means of actively moving data is a core function of a network.

A detail of the communication system that needs to be considered at the design stage is latency. An exceptionally simple system will pass messages along instantly. In simulations that don't require finite quantities of bandwidth to be taken into account (either as a parameter, or for use as the output information of the system), this is acceptable. If this is required, then packets must be able to remain 'on the wire' over time, requiring the communication system to be able to hold packet data, and serve as temporary data storage.

One consideration for this component is which form the communication should take place: direct message passing, or intelligent network-based message passing. This is expanded on further in Section II.C.1.

B.3. Scheduler

In order to represent actions occuring in a linear order across the breadth of the entire simulation, a single global timer is needed. This can take the form of a single global discrete numerical value that increments once all actions that are associated with it's value are completed. Associated with this time is a scheduler, capable of holding actions that represent the changes that act upon the state components (and possibly communication system). This scheduler should take the form of an ordered list, holding all actions that should occur at a scheduled time in the location in the structure associated with the current time. Events occuring at the same discrete time are executed in arbitrary order.

The scheduler in a discrete time simulation works in discrete time 'ticks', executing all instructions that are associated with a single time tick in a non-definable order.

B.4. Execution Engine

The execution engine accesses the actions stored in the global scheduler for a time tick specified in the global clock,

and executes the associated code. It needs read and write access to all state and communication components in order to be able to change, add or remove elements. The execution engine should be able to react to events (such as the arrival of a packet), as well as simulating a 'lag' on events (such as a packet being placed on to a network arriving elsewhere on the network some time later).

Simple, common events include addition and removal of state nodes in the system, connection and disconnection of communication between two of nodes, the construction and sending of packets of data between nodes, over the communication system, and the reception and interpretation of packets of data from a particular node. These simple events can then be combined, and miscellaneous system events added, to create complex events. For instance, most simulation begins in an 'empty' state, with no nodes or communication system in place. The first event that occurs in the simulation is the addition of all nodes in the system, with the associated communication media.

Similarly, the final event in the system is typically a 'finish' event, which writes the system state into a user-readable state (which is detailed in Section V.C), and then orders the scheduler to stop execution.

Because the timing system is discrete, 'pause' options become possible by stopping execution at the end of a specific time cycle. Similarly, state saving and restoring is possible by loading the state of the system and resuming execution. Finally, though it is out of the scope of the document specified in Section I.B, a further benefit of the discrete nature of the system is the ease by which it can be divided for parallel or grid computing.

B.5. Mechanically-Focused Simulation Review

Simulations must be capable of holding a large state component, with come form of communication system that allows packets to be transferred from one node to another. These packets are sent as a result of actions that have been scheduled to occur at the time specified by the global timer, and the act of their transfer is executed by the execution engine.

C. Optional Requirements

While the components shown in Sections II.A and II.B are necessary for the operation of a network simulator as described here, the following components allow for increased detail in simulations. When correctly implemented, the accuracy of the represented simulacra to their 'real-life' counterparts is much higher.

C.1. Connectivity Maps

Simulations require some form of connectivity map to represent unreachable states in the network, default routes, etc. To properly represent an IPv4 network, some form of connectivity map is required that will allow users to specify the type of network they want designed (e.g. mesh-connected, tree-connected, etc.), in a logical or in a physical sense.

This should also allow for concepts such as firewalls and DMZs to be represented, where every host is not directly connected to the Internet, but is connected via a gateway that can filter the networks input and output.

C.2. Semantic Node Associations

Nodes need to be able to closely represent a network that exists in the real world in order to adequately simulate such a scenario. In order to do so, they must have a strong semantic association with such nodes. If the necessity of the simulation is for a specific network to be simulated, then it is possible to input directly the network structure and the semantics of those hosts.

If, on the other hand, the intended purpose for the simulation is for a larger scale network to be used, then it becomes necessary to use statistics and a random generator to create simulated networks. In this case, we can acquire information about the composition of the internet (such as the number of actual hosts, the typical structure of a network, regularity of observed operating systems and web browsers, etc.) from websites such as CAIDA. With this information, we can create a random network generator that intelligently creates networks which follow the trends of the real world.

III. Implementation details

Once the components have been decided, the program design must be solidified. However, before design can take place, several concerns need to be considered.

Because network simulations often deal with a large set of states, it is often necessary to include structures and algorithms that are more complex than will immediately appear necessary. Even class A and B networks can be of large enough scale to result in impractically slow execution of the simulator. These challenges in the areas of optimization should be considered (if not necessarily implemented) before construction of the system begins, as the solution will often require changing core components of the program.

A. Challenges

In building a simulator of any significant scale, three major challenges arise. The massive amount of memory required to hold the information pertaining to the state components becomes untenable on a single computer. This is aggravated by typically poor means of associating semantic information with nodes. Finally, access speed is an optimization challenge, due to the massive number of hosts and the need to rapidly acquire a specific host upon which to operate.

A.1 Memory Shortage

Memory shortage is a great challenge in a simulation system. If we were to try to simulate the entire Internet, we would require 2^{32} = 4294967296 hosts to be simulated (Removing Class D and E networks, it still numbers approximately four billion addresses).

If we consider the amount of memory required to hold a pointer to a node structure, we will find that the simulation will require 4294967296 x 32 = 137438953472 or roughly 140 Gigabytes of memory. This is not perfectly accurate, as special networks, such as class D and E networks which will not have to be simulated, have been ignored. However, these networks are relatively small, and the amount of

memory used to represent them would be insignificant - the amount of memory needed is still vastly greater than most modern desktop computers can hold. This does not include information about the hosts, this is merely the memory used to hold pointers to all of them. This also does not include the associated communication system which grows exponentially as we add hosts to the system.

A.2 Host Bloat

As the simulator becomes more robust, the hosts are expected to hold more and more information. If we are allowing several different concepts to be represented in our nodes, then they can quite easily hold dozens of variables defining operating system, hardware specifications, etc. As shown in Section A.1, the number of hosts in the system might be large, and the memory structures used to hold them explode the amount of memory used. The challenge of holding large quantities of information in a very large number of structures is in optimizing the node detail access, and the solution is presented in Section B.3.

A.3 Node Access Time

Most modern programming languages use a simple data structure such as a 'Vector', or 'List'. It would seem tempting to use a structure like this to store the large number of hosts in our system, as it is a standard in the language, as well as straight-forward.

The challenge arises in the means of access.

An optimization is possible if we assume that we are simulating a series of nodes from an IPv4 network. In order to extract a node from the data structure given the IP address in some form, we need to traverse the structure, comparing every element with the associated IP address (assuming that the IP address is stored as a detail in the node).

This is barely noticeable in trivial simulation examples, but when the simulation grows large, at every time tick it could quite possibly be searching through several million elements in the array, thousands of times. The access time for finding a node must be incredibly quick in order to facilitate rapid simulation.

B. Solutions

B.1 Access Optimizations

In order to solve the access time challenge presented in Section A.3, it is necessary to use optimized methods of searching for an object in memory. By using data structures that are ordered, we can improve search times significantly. Presented here are the two recommended methods: trees and hash tables.

Trees Trees can be used to traverse the IP space very rapidly and efficiently. By separating the hosts by IP address into their hexadecimal pairs, we create a tree that is four levels deep, and the closely approximates network structures. It is also easier to optimize for space, detailed in Section B.2. Because the time required to find a node is constant (four traversals), efficiency in finding nodes is markedly improved.

Hash Tables Hash tables can be used for even faster node retrieval. While trees require four traversals, hash tables can immediately return the node. The only processing in order to access the node is to apply the hashing function to the IP address of the host

B.2 Space Optimizations

A solution to the problem of memory is to only instantiate nodes and create pointers to them when they are required. When the simulation is started, the Internet can be described as a single entity. As a specific host (say, 146.231.115.89) is addressed, a new node can be dynamically created to represent the host.

At runtime, then, the host can be instantiated and detailed using statistics and randomly generated values to represent its attributes. Research into the area of statistically generating the nature of a node and it's connectivity can be found in [3].

A greater challenge exists if the user wishes to use every host on the Internet (or a particularly large network) for their simulation. In this case, it is possible to do manual page swapping to a hard-drive, though access time will be slow. Another alternative is heavy reliance on detail optimization, detailed below.

B.3 Detail Optimizations

The solution to the challenge of host memory use is to optimize the way in which memory is stored. This can be done in three ways: by altering the way in which variables are being stored in memory, by using dynamic data structures that only use memory when required, and by creating reverse detail lists.

Efficient Storage When designing nodes, this factor should be taken into account. Using large, memory inefficient data structures for these details will result in a large expenditure of memory. Avoid the use of strings and other list-based structures if possible, and prefer integers, enumerations and Boolean values. In places where strings are required, determine if it is possible to use a hash-function and simply perform comparisons.

Dynamic Data Structures It is preferable, when creating nodes, to assign no details to them, and create a dynamic data structure that can hold information that shows differences between the node and the norm.

In the case of a simulation where most nodes are heterogeneous, this may present problems. If every node differs from the norm, or if the number of details that must be represented are few, then the overhead of a dynamic data structure may overshadow the saved memory. In this case, it is better to specify details for every node statically.

Reverse Detail Lists When the number of nodes that have a specific detail is small, or, more importantly, if some action needs to performed on all nodes with a certain property, then reverse lists become a necessity. A reverse detail list stores a list of all hosts for which a detail is pertinent. An example of it's use is in Internet worm simulation: a list storing all nodes that have been infected by the worm (and updated every time the worm infects a further host) has significant benefits over storing the details of the infection in the node itself. The first of these benefits is greatly improved efficiency. In order to find all the nodes in the system that are infected with a worm, it would be necessary

to visit every node and determine it's infection status. With reverse lists, it is a simple matter to traverse the list and act upon each entry.
The exchange for efficiency in reverse detail lists is the extra memory required. The overhead of lists for details may be greater than the amount of extra memory that would be used to hold those details, especially if dynamic data structures (mentioned above) are used.

IV. IMPLICATIONS

Simulation is a common technique in research. Network simulation, specifically, is valuable as it allows the difficult and expensive tasks in network construction to be tested before any decisions are made.
By constructing robust network simulations, a simulacrum of a proposed network system can be constructed, and if found to be undesirable, the simulator can easily be adjusted to make use of or remove components of the network or to change parameters of the system.
For instance, if a large corporation or organisation were to simulate the implementation and roll-out of IPv6 as the standard layer 3 protocol for their organisation, and the simulator showed problems due to higher-level protocols, then a robust simulator could easily be altered to simulate the implementation of IPv6 in a specific section of the organisation where it would still be useful and would not result in unnecessary problems.

V. FURTHER IMPROVEMENTS

Further improvements can be added to the simulation system here. This paper has provided outlines and advice for generating a simulator, but many extensions can be added, and many optimizations can still be made. As pointed out in Section I.B, by narrowly defining scope, it is possible to recommend domain-specific improvements, and this paper has aimed to give recommendations for the generic case of discrete-time simulations.

A. Optimizations

Ethernet and TCP/IP are currently the predominant set of protocols for networking. With these in mind, optimizations can be applied that focus on their operation.
As was shown in Section III.B.1, IP addresses are one means of optimization. By carefully selecting the form in which addresses will be used (for instance, 32-bit integers, strings, an array of bytes, or some other special method), optimizations can be added to the system.
Use of secondary servers (especially an external database for holding node details and connectivity information) can often allow a simulation to focus on the execution details, while the database is used as the core for holding information.

B. Details

By knowing the type of network that is being used, more accurate networks can be detailed. Through the knowledge of underlying protocols and physical structures, such as the US cable networks, common network structures, and the accumulation of servers into a logically-seperate, physically-close space, it is possible to derive simulations that a more generic simulator would be incapable of.

C. State Saving

Due to the stateful nature of the simulation, it is simple to save the state of a network (and the seed used to get it to that state), in order to resume simulation at a later date. This can prove useful in scenarios where execution of a simulation can be terminated, or if minor adjustments can be made to the simulation after it has begun.
This will require two components to be added to the simulator: a state saving component, and a state restoring component. Due to the prevalence of text-based formats such as XML, and the ease with which a simulation can be edited in text, it is recommended that output to a file is done textually.
It is quite possible to design a network to be simulated, and save it's state at time 0. This will allow a user to easily change the network, or parameters of the simulation, and consistently use the same base network to determine the effect that this will have.

VI. FINAL OVERVIEW

Network simulation is a challenging task, with many possible problems arising in implementation, such as extreme memory requirements (Section II.A.1), aggravated by large memory footprints used per node (Section III.A.2). By implementing the robust system detailed in Section II, and applying the optimizations and solutions shown in Section III, it is possible to construct large-scale, semantically rich simulations that operate on a single, standard host.

REFERENCES

[1] V. Paxson and S. Floyd, "Why We Don't Know How to Simulate the Internet", *Winter Simulation Conference,* pp. 1037-1044, 1997.

[2] M. I. Sharif, G. F. Riley and W. Lee, "Comparative Study between Analytical and Packet-Level Worm Simulations", *Proceedings of the 19th Workshop of Advanced and Distributed Simulation,* pp. 88-98, 2005.

[3] S. Zhou and R. J. Mondragon, "Towards Modelling the Internet Topology – The Interactive Growth Model", *Proceedings of the 18th International Teletraffic Congress, Elsevier's Teletraffic Science and Engineering series*, vol. 5a, p 121, 2003.

[4] J. Leguay, *Analysis of the Internet Topology,* Thesis for Linkoping University, Paris, 2004.

[5] S. Savage, T. Anderson, A. Aggarwal, D. Becker, N. Cardwell, A. Collins, et al., "Detour: Informed Internet Routing and Transport", *IEEE Micro*, vol. 19, no. 1, pp. 50-59, Jan/Feb, 1999

[6] F. Kelly, "Mathematical Modelling of the Internet", *Proceedings of 4th International Congress on Industrial and Applied Mathematics*, pp. 105-116, 1999.

[7] D. M. Rao and P. A. Wilsey, "Multi-Resolution Network Simulations using Dynamic Component Substitution", *Proceedings of the 9th International Symposium in Modelling, Analysis and Simulation of Computer and Telecommunication Systems,* pg 142, 2001.

A Study on Enhanced Multipath Routing Protocol in Hybrid Wireless Mesh Network

JoonYoung Cho, EunHae Kim and JooSeok Song*
Department of Computer Science
Yonsei University, Seoul, Republic of Korea

***Abstract*-Hybrid Wireless Mesh Networks (HWMNs) based on the IEEE 802.11 are being increasingly deployed as a varied network of Mobile Ad-hoc Network (MANET) to provide users with extended coverage for wireless Internet access. HWAN, however, has a problem of too wide broadcasting of the Route Request (RREQ) in the route discovery of the Ad hoc On-demand Distance Vector (AODV) protocol. In this paper, to enhance the performance of the AODV protocol, we have transformed it into a multipath protocol and proposed a new protocol that overcomes disadvantage of AODV by composing the neighbor table for overloaded RREQ. The neighbor table is provided in the form of metric using several items including the bandwidth. This can bring the localization effect of the nodes and we can expect a performance improvement by preventing an overloaded RREQ. Through the simulation, we could confirm better performance in packet loss rate, throughput, packet delivery ratio and routing overhead than the DSR protocol and AODV protocol.**

I. INTRODUCTION

The wireless mesh network is based on the IEEE 802.11 and a modified form of the mobile ad-hoc network (MANET). Therefore, AP and station is ad-hoc type which the message relay is possible for other APs. This is to construct the self-configuring system and expand coverage of AP by increasing bandwidth usage. Also, the wireless mesh network has another advantage which is the compatibility to cheaper devices that are based on the IEEE 802.11 standard. Mesh network topology also provides diverse possibilities that facilitate the arrangement of public access wireless networks and ubiquitous connection environment which enables high quality application and services.

The wireless mesh network is classified into three main types according to its structure. The three main types are infrastructure mesh network, client mesh network and hybrid mesh network [1].

1) Infrastructure Mesh Network: This network is made up with static mesh routers, which have ad-hoc characteristics. The mesh routers in the infrastructure mesh network commonly play a role as a gateway towards wired network and provide complete WAN connectivity to the wireless mesh network [2].

2) Client Mesh Network: It is pure mobile ad-hoc wireless network that has independent routers, which is called mesh clients, are principally spread widely. In client mesh network structure, the network is only consisted of mobile client device [3].

3) Hybrid Mesh Network: This is the most comprehensive and interesting version of mesh network that is composed by combining the existing infrastructure mesh network structure and client mesh network structure. The mesh router exists on the hybrid wireless mesh network functions as follows and it is the biggest difference to the mobile ad-hoc network. It works as a bridge that connects different networks and provides links to the internet like gateway.

Until now, the protocol study in the mobile ad-hoc network environment has been conducted far more than the study of protocols appropriate for the hybrid wireless mesh network. Therefore, to propose an appropriate protocol for the hybrid wireless mesh network, we will propose a routing protocol that is appropriate to the new environment by transforming the routing protocol in the preceding mobile ad-hoc network to propose the appropriate protocol for the hybrid wireless mesh network. In this study, we intend to give several forms of transformation on the AODV routing protocol to fit the characteristics of the hybrid wireless mesh network environment [4]. To propose an enhanced multi path routing protocol, we will focus on the fact that among the characteristics of the mobile node in the hybrid wireless mesh network, there hardly is no mobility and the fact that in composing the points and networks, it is more complex structure than the web-formed wireless mobile ad-hoc network, not a tree-form. Also, we should keep in mind that the mesh routers are connected with the high bandwidth to support the multi destination is one of the important features of the hybrid wireless mesh network in the proposed protocol.

The remaining contents of the paper are composed as follows. In Section II, we will introduce the technical research about routing protocols. In section III, we will suggest a multipath routing protocol that is appropriate for hybrid wireless mesh network. In Section IV, through computer simulation of proposed routing protocol and existing routing protocol, we will compare the performances of proposed protocol with AODV and DSR. Finally, Section V concludes the paper.

II. RELATED ROUTING PROTOCOLS

A routing protocol is necessary for inspecting and maintaining the route path from source node to destination node. The existing ad-hoc routing protocol can be categorized into 4 types [5], [6]: on-demand (or reactive) routing protocol, table-driven (or proactive) routing protocol,

* Corresponding author.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 395–400.

and the combination of these two types, hybrid routing protocol and geographic routing protocol which is protocol and geographic routing protocol which is based on the geographic information [7], [8].

1) On-demand (or reactive) Routing Protocol: This protocol decides the routing path by sending the data packet. Therefore, it is more scalable protocol for a great size network with mobility. When the route to the destination is required, the source node initiates the route discovery to find the best route.

2) Table-Driven (or proactive) Routing Protocol: This is protocol maintains the route path by updating the routing table information regularly with no relation to the data traffic. Each node manages the routing table that has the information of each node toward the all nodes.

3) Hybrid Routing Protocol: This protocol combines the advantages of on-demand method and table-driven method. In ad-hoc network environment, the communication between the close nodes happens frequently and the further nodes use the idea that has less frequency of the communication.

4) Geographic Routing Protocol: Using the geographic information of the objective node that is provided by the Global Position System (GPS), it reduces the overhead of the routing protocol message and save the time for route discovery by limiting the range of broadcasting neighbor node.

A. Dynamic Source Routing (DSR) Protocol

Dynamic Source Routing Protocol (DSR) is a simple and efficient routing protocol that is specially designed to be used in the multi-hop wireless ad-hoc network of the mobile nodes [9]. DSR is based on the source routing method and all the nodes maintain the route cash. As you can tell from the name, the DSR protocol uses the IP source routing. All the data packets sent using the DSR protocol includes a complete list of the nodes that the packet needs to move.

During the discovery of the route, the source node broad casts the RREQ packet with a unique identification number. The RREQ packet includes the address of target node in the pre-defined route.

B. Ad hoc On-demand Distance Vector (AODV) Protocol

AODV routing protocol is responsible for the installation and maintenance of the route path of the mobile nodes that participates in the ad-hoc network and support the multi-hop routing [10]. It protects the routing loop using the destination progressive number and uses a similar route discovery process as DSR. The AODV allows the route information to be applied faster even though the destination node is newly set. AODV easily adapts to the route information in case of discovering the link breakage or change of network topology. AODV ensures the loop-free even though the topology of the ad-hoc network changes by avoiding the “counting to infinity” problem of Bellamn-Ford.

C. Requirements of Multipath Protocol

1) Loop-freedom: One of the two conditions required for conducting the multipath protocol is loop-freedom. This means that the algorithm which prevents the looping in deciding route path should be included. For instance, when the route path that transmits from node A to node B is confirmed and the route path transmitting from node B to node C is decided, the loop occurs when the node C directs towards node A. To prevent this faulty route path in advance, the sequence number is used.

2) Disjointedness: In multipath routing protocol, various paths, not just one path from the source node to the destination node is decided. At this time, the lack of commonalities between the determined paths is disjointedness. Disjointedness is classified into two types. One is node disjointedness which there is no node that shares various route paths. The other is link disjointedness which although there is a node that shares various route paths, there is no case that uses an identical link.

III. PROPOSED MULTIPATH ROUTING PROTOCOL

A. Key Idea of Proposed Routing Protocol

AODV is a very basic routing protocol that is very widely used in many networks consisted of mobile nodes. Owing to its simple but strong performance, numerous studies have been done and introduced many transformed routing protocols. However, in the case of transformed routing protocol that is based on the AODV, since it is commonly the on-demand routing protocol, there are problems with the complex networks such as the hybrid wireless mesh network. That is, in the route discovery process that requires the route path, the source node broadcasts the Route Request (RREQ) and waits for the Route Reply (RREP). At the same time, the RREQ is flooded into the all connected nodes which will flood the RREQ again until it finds the destination node. If the distance to the destination node is quite long, the flooded RREQ increase exponentially which result in a huge overhead throughout almost all of the networks.

Fig. 1 is an easy explanation of the proposed routing protocol. In the top route path in the diagram, the number of hops is smaller than below; however, the intermediate node is overloaded to reduce the efficiency of the whole link. However, on the below picture, although the number of hops has been increased, the load of the node is decreased mostly and above all. To the routing protocol that applied this idea, we can expect a reduction of overhead through the load balancing of the entire network and throughput of the route path end-to-end delay by applying the multipath routing technique which is getting popular these days.

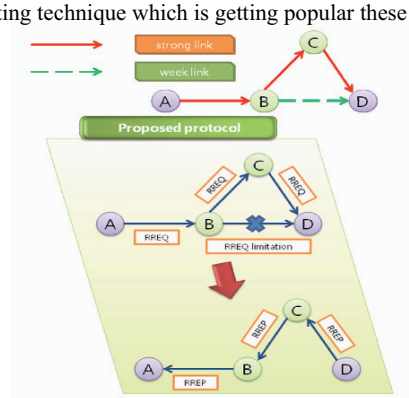


Figure 1. Key idea of proposed routing protocol.

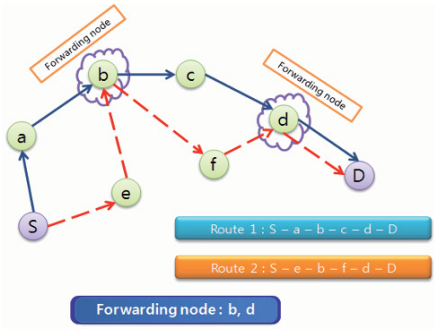


Figure 2. An example of the forwarding node.

B. *Forwarding Node*

When the route discovery of the given routing protocol is completed and the route path is decided, we will define the node that is conducting the forwarding as a forwarding node. As various working route paths use the identical forwarding node repeatedly, the route paths get firmer. On the contrary, when the various route paths are composed of the route path with minimized number of forwarding node, these route paths get enhanced stability to the link breakage of the network and bring the load balancing effect. To help explaining the forwarding node, the following Fig. 2 is an example of the forwarding node.

If there is no forwarding node, the routing protocol makes all the node disjoint. In this study, by providing a parameter that is affected by the forwarding node on the metric, we will propose an algorithm for calculating the appropriate routing path for the network. That is for the environment that requires a strong network, it gives points for increasing the number of forwarding nodes and for the environment that requires stable and efficient network, it makes to lose points for the increase of number of forwarding nodes. By controlling the forwarding node with transforming the metric, disjointedness is controlled in deciding the multipath. A detailed discussion on this method will be continued on the next part.

C. *Neighbor Table Metric*

A node saves the metric information of neighbor in the neighbor table and uses the metric as information for preventing the indiscreet broadcasting of the RREQ in the route discovery process. The AODV based routing protocol follows the hop by hop method. However, in this paper, we will propose the limitation of RREQ during the route discovery process using the metric information based on the bandwidth and make more efficient information transfer possible by profiting from the bandwidth even though the hop gets longer by using the metric of bandwidths. The definition of the variables required for explaining the metric is as the following Table I.

Metric is consisted of above variables and calculated through the following formulas. NTM refers Neighbor Table Metric.

If the node *i* is a forwarding node,

$$NTM_i = (\sum Bandwidth_{ij} * (1 - LER_{ij}) - CTV_i - \sum FTV_i) * \alpha \quad (1)$$

If the node *i* is not a forwarding node,

$$NTM_i = (\sum Bandwidth_{ij} * (1 - LER_{ij}) - CTV_i) * (1 - \alpha) \quad (2)$$

TABLE I
The components of neighbor table metric

Metric	Description
Distinct rate (α)	A parameter that decides whether to load the point or penalty for the forwarding node. (0 ≤ α ≤ 1) - To increase the number of forwarding nodes, set α value close to 0. - To reduce the number of forwarding nodes, set α value close to 1.
$Bandwidth_{ij}$	Bandwidth of the link directed from node I to node j.
LER_{ij}	Link Error Rate; The error rate of the link directed from node i to node j.
CTV_i	Current Traffic Volume; A current processing traffic volume of node i.
FTV_i	Forwarding Traffic Volume; A traffic volume that forwards to node i.

If you look at the metric, it uses the bandwidth. In fact, in the wireless network environment, the problem of getting an accurate bandwidth is very difficult. Because of the competition and diminution of the devices that supports each other, there are many cases that does not support as much as the supporting bandwidth. However, in case of this metric, the fact that it doesn't require an accurate bandwidth needs to be attended. It is because, this metric information is only used as the information that simply limits the flooding of the RREQ. Eventually, the proposed routing protocol basically supports the hob by hop and with the metric that is based on the bandwidth, only the broadcasting of RREQ is limited. The reason for multiplying bandwidth with the value which subtracted the LER from 1 is that if the error rate of the link is 30%, even though the bandwidth is 100, the amount which the real bandwidth available for communication is 100 x 30% = 70. We get the sum of these values to add up all the bandwidth in case there is more than one possible channel. Also, it should be excluded from the bandwidth as much as the amount of the traffic that the current node is using and for the forwarding node, we subtract the total amount of the traffic amount used for forwarding. Lastly, according to α, which is a parameter for loading the point or a penalty to the disjoint for the forwarding node, it needs to be multiplied as much as the weight value α and for the contrast, it needs to multiply the value which subtracted α from 1 to change the ratio.

D. *Route Discovery in Proposed Protocol*

The links to the bad metric node on the source node is ignored and the RREQ is broadcasted to the good value. On each node that the RREQ has been received, similarly, the same process is conducted through the metric values of the neighbor table. At last, when the RREQ is received to the

destination node, the RREP from the node to the source node is returned. To support the multipath, the RREQ reception of the destination node allows the repetition and the RREP is sent to each route ID. This RREP is re-transmitted to the source node to decide the route path and the data transmission begins. The characteristics of the decided route path rely on the parameters that decide the metric values.

IV. PERFORMANCE EVALUATION

A. *Simulation Environment*

To show the performance improvement through the measuring the performance of the proposed routing protocol and comparing the performance to the existing routing protocol, we conducted a computer simulation. Network Simulator 2 (NS-2) [11] with CMU's wireless extension and C++ Language were used for the simulation program. The NS-2 is a discrete event simulator for studying the network performance such as TCP, and routing protocols in cable and wireless network environment developed in University of California at Berkeley and VINT Project. The C++ language was programmed on the console mode on the Visual C++ 6.0 of the Microsoft.

Various parameters of the simulation is shown in the Table II. In the experiment, the standardized 802.11b radio was applied and the channel rate was set as 2Mbps. The transmission radius was 250m and career sensing range was 550m. The transmission of each data packet is pre-administrated with the Request-To-Send / Clear-To-Send (RTS/CTS) handshake on the MAC layer. The total simulation time is 700 seconds and the interface queue is 50-packet drop-tail priority queue.

The one that is used as a traffic type in the simulation is the Constant Bit Rate (CBR) data session. The pair of the source node and destination node is selected randomly among all the nodes. The CBR session in all the scenarios have packet size of 512 byte and each packet is transmitted at the rate of 4 packets per second.

TABLE II
Network environment for computer simulation

Simulation Time	700s
Simulation Area	1000m x 1000m
Number of Mesh Routers	36
Number of Mesh Clients	64
Movement Model	Random way point
Channel Rate	2 Mbps
Transmission Range	250 m
Career Sensing Range	550 m
Client Speed	0 ~ 20 m/s
Traffic Type	CBR (UDP)
Packet Size	512 byte
Packet Rate	4 packet/s

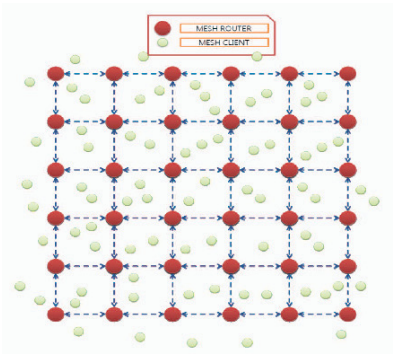


Figure 3. Network topology of simulation scenario

B. *Simulation Scenarios*

In the mesh network that covers 1 square kilometer geography, 36 mesh routers in total are installed. The topology is composed as shown in the Fig. 3 from no. 1 to 64. This router handles 20 connections between the allocated mobile clients from number 1 to 64 simultaneously. The source and destination of these connections are selected in between the mesh clients. Each node uses the random way point model. This is the location of the destination which the mesh client halts for a while and moves to the final destination location in random speed of 0~20m/s. When the transmission is completed, it stops for a while and moves to other random location. All RREQ packets use un-slotted Carrier Sense Multiple Access protocol with Collision Avoidance (CSMA/CA) for broadcasting. Each reception node in CSMA/CA waits for free channels. If the channel is free, the node holds on to the channel and sends the data. In the case of collision, the collided station stops transmission and waits while using the binary exponential back-off algorithm. In unicast packet, the node sends the Ready-to-Send (RTS) first to reserve the channel. The scheduled reception node is responsible for sending the Clear-to-Send (CTS) frame to the sender. All the nodes with overloaded RTS and CTS frame stops transmitting during the Network Allocation Vector (NAV). After receiving CTS, the packet is transmitted [12].

C. *Simulation Results*

The following Fig. 4 shows the packet loss rate in the simulation environment and the scenario. The following Fig. 5 shows the aggregate throughput in the simulation environment and the scenario. Fig. 6 shows the packet delivery ratio in the simulation environment and the scenario. Compared to the existing routing protocol, the proposed routing protocol shows the improved performance at the ratio of number of successful packets to the overall packets. Already, we have looked at the result of the packet loss and the throughput at the previous part. Through this result, we can calculate the packet delivery ratio. The proposed routing protocol shows higher packet transmission ratio than the existing DSR and AODV protocol. When there is no mobility, the proposed routing protocol shows a high packet transmission ratio of 95.02%. This is because there is less packet loss and higher throughput. Also for the packet transmission ratio, similar to the packet loss and the throughput, when the mobility increases to 20m/s, the

performance of the overall packet transmission ratio is reduced to 89.52%. However, compared to DSR and AODV, especially for DSR, the level of overall performance reduction is lesser. This is a result that the proposed routing protocol supports the alternative path for lesser creation of the link breakage. The following Fig. 7 shows the routing packet overhead in the simulation environment and the scenario. The proposed routing protocol limits the RREQ, thus it shows lesser routing overhead than the AODV. However, opposite to the protocol that is based on the AODV, the DSR protocol shows the efficiency in the unique caching strategy that limits the control packet overhead. Using various mechanisms learned while this protocol is forwarding, the information is accumulated on the cache. As the mobility increases, the link breakage increases and since the search for new route is required, the routing overhead of overall routing protocol increases.

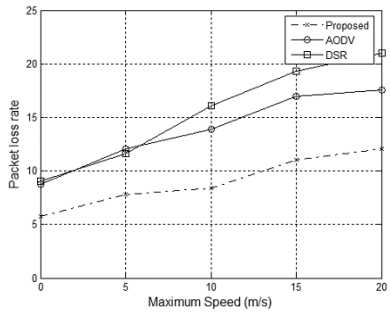


Figure 4. The simulation result of packet loss rate.

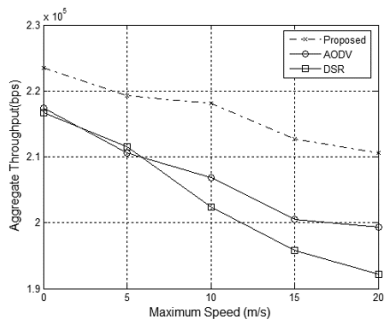


Figure 5. The simulation result of aggregate throughput.

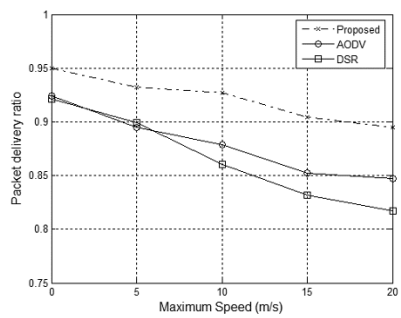


Figure 6. The simulation result of packet delivery ratio.

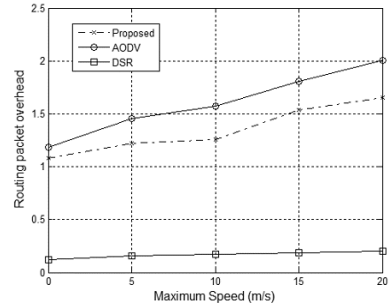


Figure 7. The simulation result of routing packet overhead.

V. CONCLUSION

Recently, the hybrid wireless mesh network is getting attention. The mesh router works as the gateway to the wired-network and it provides WAN connectivity and the mesh clients participates on the technical functioning of the network as a mobile device. In addition, the mesh router supports the multi destination connected to other mesh routers with high bandwidth. In this paper, we proposed a new multipath routing protocol for hybrid wireless mesh network. We have transformed it into a multipath routing protocol and proposed a protocol that overcomes by composing the neighbor table for overloaded RREQ. The neighbor table is provided in the form of metric using several items including the bandwidth. It required a parameter for limiting the RREQ. Through the computer simulation, we could confirm improvement in packet loss rate, throughput, packet delivery ratio and routing overhead. This seems to be a result of dispersion of the load to various route paths due to the transformation into the multipath. However, the proposed protocol showed decrease of optimality of the path compared to the existing AODV routing protocol. However, since it is less dependent on the size of the net, it showed strength in terms of the scalability. With a current algorithm, the quality of service, security and energy efficiency was not a consideration. In the future works, there should be studies of increasing the path optimality and more efficient methods for other requests which is expected to be provided in the wireless mesh network to enhance the function of the protocol.

ACKNOWLEDGMENT

This work was supported by the Korea Science and Engineering Foundation(KOSEF) grant funded by the Korea government(MOST) (No. R01-2006-000-10614-0).

REFERENCES

[1] I. F. Akyildiz and X. Wang, "A Survey on Wireless Mesh Networks," IEEE Communications Magazine, vol. 43, no. 9, pp. S23-S30, 2005.
[2] IEEE, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications 802.11," 1997.
[3] A. A. Pirzada and C. McDonald, "Establishing Trust in Pure Ad-hoc Networks," in Proceedings of the 27th Australasian Computer Science Conference (ACSC), vol. 26. Australian Computer Society, 2004, pp. 47-54.
[4] K. Ramachandran, M. Buddhikot, G. Chandranmenon, S. Miller, E. Belding-Royer, and K. Almeroth, "On the Design and Implementation of Infrastructure Mesh Networks," in Proceedings of the IEEE Workshop on Wireless Mesh Networks (WiMesh). IEEE Press, 2005.
[5] E. M. Royer and C. K. Toh, "A Review of Current Routing Protocols for Ad hoc Mobile Wireless Networks," IEEE Personal Communications Magazine, vol. 6, no. 2, pp. 46-55, 1999.
[6] Mohammad I. . et. al.,"The hand book of Ad hoc Wireless Networks", CRC Press, 2003.
[7] B. Karp and H. T. Kung, GPSR: Greedy perimeter stateless routing for wireless networks, Proceedings of ACM Mobi-Com'00, Aug. 2000.
[8] I. S. P. Bose, P. Morin and J. Urrutia, Routing with guaranteed delivery in ad hoc wireless networks, ACM/Kluwer Wireless Networks, November 2001.
[9] D. B. Johnson, D. A. Maltz, and Y. Hu, "The Dynamic Source Routing Protocol for Mobile Ad hoc Networks (DSR)," IETF MANET, Internet Draft, 2003.
[10] C. Perkins, E. M. Royer, and S. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing," IETF RFC 3561, 2003.
[11] NS, "The Network Simulator," http://www.isi.edu/nsnam/ns/, 1989.
[12] A. S. Tanenbaum, Computer Networks, 4th ed. Prentice Hall, 2002.

Pseudorandom Number Generation Using Cellular Automata

Byung-Heon Kang, Dong-Ho Lee, and Chun-Pyo Hong
Dept. of Computer and Communication Engineering, Daegu University, Kyungsan, 712-714, Korea
bhkang1970@hotmail.com, dhlee@dsp.daegu.ac.kr, cphong@daegu.ac.kr

***Abstract*-High-performance pseudorandom number generators (PRNGs) play an important role in a variety of applications like computer simulations, and industrial applications including cryptography. High-quality PRNG can be constructed by employing cellular automata (CA). Advantage of the PRNG that employs CA includes that it is fast and suitable for hardware implementation. In this paper, we propose a two-dimensional (2-D) CA based PRNG. Our scheme uses the structure of programmable CA (PCA) for improving randomness quality. Moreover, for reducing of serial correlations among the produced pseudorandom bits, a consecutive bits replacing spacing technique is proposed. Finally, we provide experimental results to verify the randomness quality using ENT and DIEHARD test suites.**

I. INTRODUCTION

The security of many cryptographic systems depends on the generation of numbers with randomness and unpredictability. For this security, random number generator is used. Most random number generators are deterministic in nature, in a sense that, random numbers are generated using mathematical methods. Mathematical methods are available to prevent wrong simulation results caused by inappropriate pseudorandom number generator (PRNG). Statistical tests are conducted to ensure a PRNG produces numbers that are uniformly distributed, uncorrelated with extreme long period.

There are quite a number of PRNGs. Some examples are linear feedback shift register (LFSR), and cellular automata (CA) based PRNGs. In the past decades, CA-based PRNGs were studied primarily because of its easy implementation in hardware and its parallelism [1]. The majority of research on CA-based PRNG has been focused on one-dimensional (1-D) CA with nearest three-cell neighborhood, also known as elementary CA [2]. Recently, there is a research trend such that increased complexity from hybrid CA cell configurations and increased CA dimensionality can lead to better performance. Researchers have tried to improve their results by designing new configurations with hybrid cell rules and new boundary conditions [3] also experimented with increasing the CA dimensionality to obtain better result [4]. The neighborhood for two-dimensional (2-D) CA leads to better performance is shown in [4].

A 2-D CA based PRNG is proposed in this paper. Advantage of the PRNG that employs CA includes that it is fast and suitable for hardware implementation. Our scheme uses the structure of programmable CA (PCA) with two control signals. The CA rules used our PRNG are additive rules which has XOR and XNOR logic with a hybrid dependency. Moreover, a consecutive bits replacing technique is proposed for minimizing of serial correlations among pseudorandom random bits. Finally, this paper provides experimental results to verify the randomness quality using ENT and DIEHARD test suites.

In the next Section, we introduce some background knowledge upon which the subject matter of this paper is based. In Section 3, we provide a 2-D PCA based PRNG which has a high randomness quality. Section 4 shows the results on the proposed PRNG structure. Then, our scheme is compared with previous well-known PRNG scheme in terms of an ENT and DIEHARD test suites. Finally, Section 5 provides a conclusion.

II. BACKGOUND

A. Cellular Automata

A CA is a collection of simple cells connected in a regular fashion. Wolfram [5] pioneered the investigation of CA as mathematical models for self-organizing statistical systems and suggested the use of a simple two-state, three-neighborhood (left, self and right) CA with cells arranged linearly in one dimension. The CA structure investigated by Wolfram can be viewed as a discrete lattice of cells where each cell can assume either the value 0 or 1. The next state of a cell is assumed to depend on itself and on its two neighbors (three-neighborhood dependency). The cells update in discrete time steps according to some deterministic rule that depends only on local neighbors.

For a one-dimensional (1-D) CA, a cell is connected to r local neighbors on either side, where r is referred to as the radius. Thus, each cell has $2r + 1$ neighbors. For a two-dimensional (2-D) CA, two types of cellular neighborhoods are usually considered: five cells, consisting of the cell along with its four immediate non-diagonal neighbors and itself (also known as the von Neumann neighborhood) and nine cells, consisting of the cell along with its eight surrounding neighbors and itself (also known as the Moore neighborhood). The von Neumann neighborhood type is considered only in this paper.

The randomness quality is affected by a boundary condition. In this point of view, a periodic boundary condition is expected to generate better. And, it is proved that a periodic boundary

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 401–404.

condition (that is, the extreme cells are adjacent to each other) is better than the null boundary condition (that is, the extreme cells are connected to logic 0-state) in random number generation [6]. Thus, a periodic boundary condition is considered only in this paper.

The basic objective of a programmable CA (PCA) is to configure the CA with different rules at different instants. For this objective, a PCA allows spatial and temporal variations in the state transition rules within a CA, according to some external control structure [7]. By allowing more complex control, one can introduce immense flexibility to this programmable structure. Using such a structure for all of the cells, all possible additive rule combinations can be achieved to realize any hybrid additive CA. A PCA with two external controls is used in our scheme.

B. PRNGs based on Cellular Automata

The first CA application for pseudorandom number generator (PRNG) design has been reported by Wolfram. Wolfram described random sequence generation by simple 1-D CA with three neighborhoods [8]. The work focused on the properties of rule 30. Seven statistical tests were performed and the rule 30 was shown to be superior to the LFSR in the generation of random sequences.

Recently, the Guan et al.'s PRNG is constructed from 5×10 2D-CA structure for reducing the number of cells [4]. The PRNG structure has same rules except for the first and last columns. Thus, 40 cells among 50 cells have same rules. The hybrid CA structure can lead to better performance [9]. We can find that the results of this PRNG have a low pass rate as compared with our PRNG. The time-space diagram for visual test is described in Section 4.

III. 2D-PCA PRNG

In this Section, we propose a 2-D PCA based PRNG with two external signals. It produces high-quality random number sequences to minimize serial correlations among pseudorandom bits. For better randomness quality, our PRNG uses PCA structure with two external control signals (self and sign control signal). Moreover, we adopt a consecutive bits replacing technique for reducing serial correlations among pseudorandom bits. Thus the random numbers can be generated at a much higher rate. It is shown that our PRNG produces high-quality random number sequences on well-known test suites, i.e., ENT and DIEHARD test.

A. Rule Numbering Scheme

In this Section, for delineating our PRNG, the rule numbering scheme [10] is introduced. The von Neumann neighborhood type is only considered in this paper, thus limiting the already search space size. Since there are 64 possible additive rules-those involving only XOR and XNOR logic, 6 bits are needed to describe a rule. Let $s_{i,j}(t)$ be the state of the cell at row i and column j at time step t. The state of (i, j)th cell at the next time step, $s_{i,j}(t+1)$ is then computed as follows: $s_{i,j}(t+1) = X \oplus (C \cdot s_{i,j}(t)) \oplus (N \cdot s_{i-1,j}(t)) \oplus (W \cdot s_{i,j-1}(t)) \oplus (S \cdot s_{i+1,j}(t)) \oplus (E \cdot s_{i,j+1}(t))$ where $\oplus$ and $\cdot$ denote XOR and AND operations, respectively, and X, C (centre), N (north), S (south), W (west), E (east) are binary variables. C, N, S, W, and E denote that the respective neighboring cell state is taken into account a value of 1 or 0. The binary variable X demarcates linear (X = 0) from complemented (X = 1) rules. The rule number of a cell is then given by the 6-bit string $XCNWSE$.

B. 2-D PCA PRNG

We illustrate the proposed 2-D PCA based PRNG in this Section.

In order to produce pseudorandom numbers having a proper cycle length, some of the notations used in the proposed PRNG are defined as follows:

- $P^{(0)}$ is a 64-bit initial vector loaded into the global states of a 8×8 2-D PCA.
- SI is a 64-bit initial value loaded for a self control signal.
- CI is a 64-bit initial value loaded for a sign control signal.
- R is a repeating counter for producing pseudorandom numbers of a demanded cycle length.
- G is an initial configuration counter increased by 1 until R.
- $P^{(i)}$ is a 64-bit output sequence, $1 \le i \le R$.
- $S^{(k)}$ is an output sequence where $i \bmod 4 = k$, $S^{(k\text{-}1 \bmod 4)} = P^{(i)}$, $0 \le k \le 3$, $1 \le i \le R$.
- S is a final pseudorandom numbers of a $(64 \times \mathrm{R})$-bit size.

Our PRNG processes the following three steps:

Step 1. Initialization Step: Generates three initial seeds, randomly. Then, the initial seeds $IV1$, $IV2$ and $IV3$ are stored to $P^{(0)}$, SI and CI.

Step 2. Evolving Step: Compute repeatedly $P^{(G+1)} = F(r_{i,j}, P^{(G)})$, where $0 \le G \le R$, $F()$ is called the updating function of our PRNG. The output sequence $P^{(1)}P^{(2)} \ldots P^{(R\text{-}1)}P^{(R)}$ are generated in this step.

Step 3. Consecutive Bits Replacing Technique Step: Generates a final pseudorandom numbers with reduced serial correlations.

These three steps are described in more detail as the following sections.

The Initialization Step. In an initialization step, the initial cell values (0 or 1) of 8×8 2-D PCA are determined by the initial seed $IV1$. Then the 2-D PCA structure is constructed having a different rule at each cell. The rule number of each cell is decided by the initial control signals $IV2$ and $IV3$. The initial three values for $IV1$, $IV2$ and $IV3$ are one-time values with 64 bits. The initial seeds $IV1$, $IV2$ and $IV3$ are stored to $P^{(0)}$, SI and CI for the next updating. The initialization step is performed only one time.

The Evolving Step. The main purpose of evolving step is a generation of a high-quality random number sequence as used on the PCA structure. Similar to a conventional PCA, it has a localized state transition rule for evolving a cell state. For a state transition of neighborhood, it uses the nearest five-cell neighborhood having a radius r = 1. A complexity of PRNGs based on a PCA lies on the presence of the additional rule selection signal for the purpose of dynamic rule selection. For increasing a complexity of our PCA structure, we adopt two

control signals - a self and sign control signal. The rule of a cell is determined by current two control signal values. According to the self control signal, a PCA decides whether a self-cell affects an update of the cell at the next time step or not. According to the sign control signal, a PCA decides that the result of an update is maintained or complemented. The choice of transition rules is important for a PRNG having a good randomness quality. Many studies have found that rules 31, 47 and 63 tend to produce good random bits in 2-D CA [11]. Therefore, in our PRNG, two linear rules (15, 31) and two complemented rules (47, 63) are used including the above

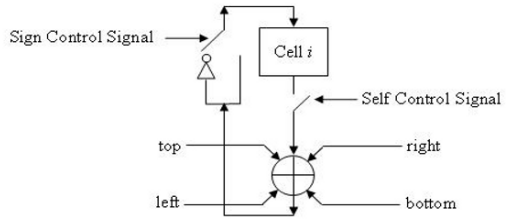


Fig. 1. PCA cell structure

mentioned rules. Fig. 1 shows a PCA cell with two control signals.

Table 1 shows the implemented rule according to the values of two control signals. This is decided by the values of two control signals.

TABLE I
IMPLEMENTED RULES

Created Rule Number	Complemented Control Signal	Self Control Signal
15 (001111)	0	0
31 (011111)	0	1
47 (101111)	1	0
63 (111111)	1	1

The structure of 2-D PCA with two control signals is shown

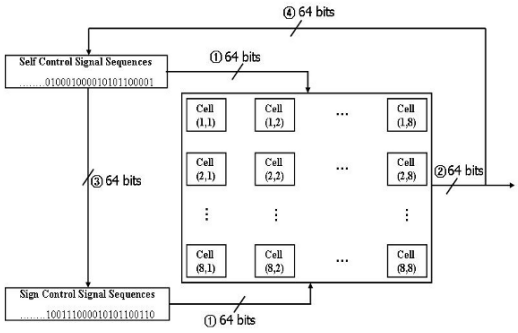


Fig. 2. 2-D PCA structure

in Fig. 2. The numbers in the circle indicate the execution sequence of 2-D PCA in Fig. 2.

Suppose that the updating step consists of processing of R rounds. The operation can considered as a transformation of $F()$. The operations in each step are done simultaneously on all bits in updating function. Let $r_{i,j}$ be the rule number of each cell at row i and column j in a 8×8 2-D PCA structure where $1 \le i, j \le 8$. The execution of updating process is as follows:

- Decides the rule $r_{i,j}$ of each cell using SI and CI.
- Generates a next state $P^{(G+1)} = F(r_{i,j}, P^{(G)})$.
- Replaces a sign control signal with a self control signal, and a self control signal with an updated state for the next updating.

The Consecutive Bits Replacing Technique Step. In order to reduce the serial correlation between bits in the pseudorandom number sequences, we carry out a max-run test. A max-run test returns the number of repeated bits (0 or 1) among the pseudorandom numbers. If the returned number is greater than 32, some bit is replaced from 0 or 1 to 1 or 0. Suppose that BP is a bit position where some bit among the repeated bits, and IC is a first bit position of pseudorandom number sequence having the same 32-bit pattern. Fig. 3 shows an example of a consecutive bits replacing technique. Now suppose that some repeating bits appear where the value of IC is 10,213. Thus, the bit of BP at the position 10.208 ($BP = IC -$

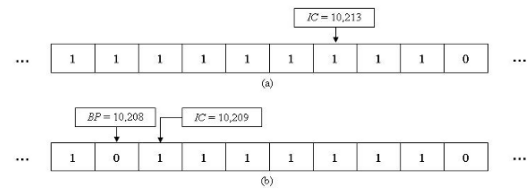


Fig. 3. Example of Step 3: (a) previous (b) next

(IC mod 32)) is replaced from 1 to 0. Then, the value of IC is changed to 10,209 ($IC = BP + 1$) for the next process.

IV. EXPERIMENTAL RESULT

In order to compare to the randomness quality schemes, Guan et al's [4] PRNG are chosen which are known as the PRNG having the best randomness quality based on 2-D CA. The experimental results are compared to the randomness quality using the ENT and DIEHARD test suites.

ENT test is a collective term for the three tests which are the Entropy, Chi-square, and Serial correlation coefficient (SCC) test. Table 2 shows each test name and max grade for a good randomness.

DIEHARD test is important because it seems to be the most powerful and difficult test suite to pass. DIEHARD test consists of 18 different tests. Results of test are called p-value. A pass is considered when the p-value is $0.025 < p < 0.975$ at the 0.05 level.

TABLE II
MAX GRADE OF ENT TEST SUITE

Test Name	Max Grade
Entropy	Close to 8.0
Chi-square	Close to 127.5
SCC	Close to 0.0

The experimental conditions are identical for two PRNGs. An identical initial seeds are used among each PRNG. The length of the tested sequences is 10 MB which is needed for

DIEHARD test. The test has performed a total of 100 experiments for comparison of the randomness quality among two PRNGs. In average *p*-values, two PRNGs passed all 18 DIEHARD tests. For more detailed analysis, we compare the pass rate. Table 3 shows the pass rate of DIEHARD test. When the tests produce *p*-values more than 2, a pass is considered when all the *p*-values are passed more than 80% at the 0.05 level.

TABLE Ⅲ
THE PASS RATES OF DIEHARD TEST SUITE

Test Name	Max Grade	
	Guan's	Ours
Birthday spacing	97%	100%
Overlapping permutation	70%	74%
Binary rank 31×31	100%	100%
Binary rank 32×32	93%	97%
Binary rank 6×8	95%	96%
Bitstream	100%	100%
OPSO	100%	100%
OQSO	23%	100%
DNA	100%	100%
Count the ones 01	92%	96%
Count the ones 02	85%	100%
Parking Lot	97%	100%
Minimum distance	98%	100%
3DS spheres	100%	98%
Squeeze	100%	95%
Overlapping sum	95%	94%
Runs	91%	93%
Craps	97%	90%

In Table 3, test results show that the randomness quality of our PRNG is superior to Guan et al.'s.

Finally, we present the 2-D time-space diagram for visual test in Fig. 4. Our PRNG has 64 cells and Guan et al.'s PRNG has 50 cells, and all schemes run for identically 200 time steps. An identical initial seeds are used to this test having the final bit 0. The other bits excluding the final bit have all 1. The *x*-axis stands for the number of cells, while the *y*-axis traces cell transitions from the time step 0 to 200. Fig. 4 (b) shows that ours does not have any repeated patterns in their diagrams. However, Guan et al.'s scheme has some regular patterns in Fig. 4 (a).

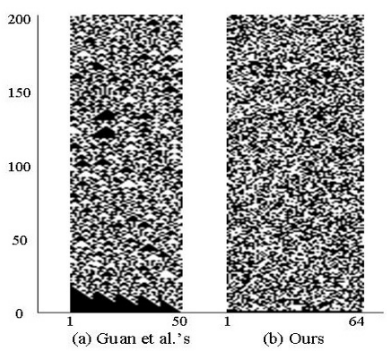


Fig. 4. Time-space Diagram

V. CONCLUSION

This paper has proposed an efficient PRNG based on 2D PCA with two control signals (a self control signal and sign control signal). We have shown how the behavior of the state transition can be produced by dynamic state transition rules via two control signals. The self and sign control signal in our scheme lead to a new hybrid neighborhood and state transition rule. Moreover, a consecutive bits replacing technique has been used in order to minimize a serial correlation.

For experimentation, ENT and DIEHARD test are used to verify the randomness quality between our PRNG and Guan et al.'s PRNG. As a result of ENT test, it is shown that the value of three tests (Entropy, Chi-square and SCC) has closed to maximum-grading. Then we have tested on DIEHARD for comparison between our PRNG and the well-known PRNG. Ours has produced a good quality of pseudorandom numbers that pass all DIEHARD tests.

Consequently, the conclusion is that the proposed PRNG based on 2D PCA satisfied the ENT and DIEHARD test and produced good quality of random numbers without a loss of an original cycle length. Thus, we expect that the proposed PRNG will be efficiently used in variety of applications including computer simulations and cryptography.

REFERENCES

[1] M. Matsumoto, "Simple Cellular Automata as Pseudorandom *m*-sequence Generators for Built-In Self-Test", *ACM Transactions on Modelling and Computer Simulation*, ACM Press, New York, 1998, Vol. 8, No. 1, pp. 31-42.

[2] S. Wolfram, *Theory and Applications of Cellular Automata: Including Selected Papers 1983-1986*, River Edge, NJ: World Scientific, Champaign, 1986.

[3] S. U. Guan and S. Zhang, "A Family of Controllable Cellular Automata for Pseudorandom Number Generation", *International Journal of Modern Physics C*, World Scientific Publishing, 2002, Vol. 13, No. 8, pp. 1047-1073.

[4] S. U. Guan, S. Zhang and M. T. Quieta, "2-D CA Variation With Asymmetric Neighborship for Pseudorandom Number Generation", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, IEEE, 2004, Vol. 23, No. 3, pp. 378-388.

[5] S. Wolfram, "Statistical Mechanics of Cellular Automata", *Review of Modern Physics*, APS physics, 1983, Vol. 55, pp. 601-644.

[6] S. Wolfram, "Cryptography with Cellular Automata", *Advances in Cryptology: Crypto '85 Proceedings*, Springer-Verlag, New York, 1985. LNCS 218, pp. 429-432.

[7] S. Nandi, B. K. kar and P. P. Chaudhuri, "Theory and Applications of Cellular Automata in Cryptography", *IEEE Transactions on Computers*, IEEE, 1994, Vol. 43, No. 12, pp. 1346-1357.

[8] S. Wolfram, "Random Sequence Generation by Cellular Automata", *Advanced in Applied Mathematics*, Academic Press, Orlando, 1986, Vol. 7, pp. 123-169.

[9] P. D. Hortensius, R. D. Mcleod, D. M. Miller and H. C. Card, "Cellular Automata-Based Pseudorandom Number Generators for Built-In Self-Test", *IEEE Transactions on Computer-Aided Design*, IEEE, 1989, Vol. 8, No. 8, pp. 842-859.

[10] D. R. Chowdhury, I. Sengupta and P. P. Chaudhuri, "A Class of Two-Dimensional Cellular Automata and Their Applications in Random Pattern Testing", *Journal of Electronic Testing: Theory and Applications*, Kluwer Academic Publishers, Norwell, 1994, Vol. 5, No. 1, pp. 67-82.

[11] M. Tomassini, M. Sipper and M. Perrenoud, "On the Generation of High Quality Random Numbers by Two-Dimensional Cellular Automata", *IEEE Transactions on Computers*, IEEE, 2000, Vol. 49, No. 10, pp. 1146-1151.

An Efficient Estimation Algorithm for MIMO OFDM System Using Turbo Codes

Khalida Noori
Comm. System Eng Dept.
National University of Sciences and Technology
Rawalpindi, Pakistan
khalida.noori@niit.edu.pk

Sami Ahmed Haider
Electrical Engineering Dept.
National University of Sciences and Technology
Rawalpindi, Pakistan
kshoab@yahoo.com

Shoab Ahmed Khan
Computer Engineering Dept.
National University of Sciences and Technology
Rawalpindi, Pakistan
kshoab@yahoo.com

***Abstract*-In this paper we apply turbo codes to multi input multi output orthogonal frequency division multiplexing (MIMO OFDM) system with recursive least square (RLS) algorithm used for channel estimation to upgrade the performance of functioning MIMO OFDM system. We use RLS algorithm because it converges faster, so it gives better results as compare to others algorithms Results are simulated over a time varying Rayleigh fading additive white Gaussian noise (AWGN) channel. End goal is to boost system performance in terms of bit error rate (BER) by adding turbo codes. MAP algorithm is used for decoding. We analysis the results for 2x2 and 3x3 transmit and receive antennas with different modulation schemes. We evaluate that there is tremendous improvement in BER by adding turbo codes with RLS algorithm.**

I. INTRODUCTION

MIMO-OFDM is regarded as a capable solution to enhance the data rate and spectrum efficiency of upcoming wireless communication systems. MIMO-OFDM combines the advantages of both MIMO system and OFDM system, i.e., enormous capacity in addition to protection against delay spread [1, 2,3].

Equalization is an effectual technique to combat ISI. Equalizer must be adaptive since the channel is generally unknown and time varying [4]. Since an adaptive equalizer compensates for an unknown and time-varying channel, it requires a specific algorithm to update the equalizer coefficients and track the channel variations. This leads to the family of powerful, complex, adaptive signal processing technique known as recursive least squares (RLS), which significantly improves the convergence of adaptive equalizers [5]. Of all practical error correction methods known to date, turbo codes come closest to approaching the Shannon limit. . Turbo codes achieve good BER at low SNR. They make it possible to increase data rate without increasing the power of a transmission, or they can be used to decrease the amount of power used to transmit at a certain data rate. [6,8]

II. TURBO CODED MIMO OFDM SYSTEM

A MIMO OFDM system transmits independent modulated data from multiple antennas simultaneously. OFDM extends directly to MIMO channels with the IFFT/FFT and CP operations being performed at each of the transmit and receive antennas. MIMO-OFDM decouples the frequency-selective MIMO channel into a set of parallel MIMO channels.

In turbo codes two Recursive Systematic Convolutional encoders are used to code the same input bits, but an interleaver is placed between the them. The outputs from the two component codes are then punctured and multiplexed. The decoder operates iteratively, and in the first iteration the first component decoder takes channel output values only, and produces a soft output as its estimate of the data bits. The soft output from the first encoder is then used as additional information for the second decoder, which uses this information along with the channel outputs to calculate its estimate of the data bits. This cycle is repeated, and as the number of iteration is increased, the Bit Error Rate (BER) of the decoded bits tends to fall. However the improvement in performance obtained with increasing numbers of iterations decreases as the number of iterations increases. Hence, for complexity reasons, usually only about 8 iterations are used [7].

Figure 1 shows system diagram of turbo coded MIMO OFDM system. The stream of input bit is turbo encoded and then modulated. The modulated data is than converted from serial to parallel and IFFT operation is performed and the symbols are modulated onto the subcarriers. To avoid intersymbol interference and mitigate the effect of delay spread cyclic prefix is added to the OFDM symbol and is transmitted. At the receiver side vice versa operation is performed.

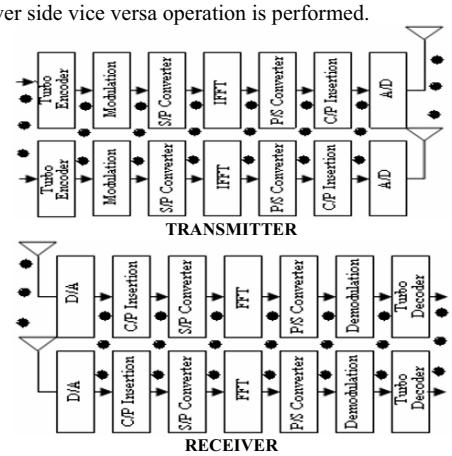


Figure 1:System diagram of turbo coded MIMO OFDM system

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 405–408.

III SIMULATION SETUP

Simulation is done for 2x2 and 3x3 Tx and Rx antennas. OFDM with 32 subcarriers and different modulation schemes are used. Recursive least square algorithm is used for channel estimation.Turbo encoder is used for coding and MAP algorithm is used for decoding. Simulation results are shown for Rayleigh fading with AWGN channel.

IV SIMULATION RESULTS

Figures shown below are BER plots of turbo coded MIMO OFDM system. Figure 2 and Figure 3 shows BER of turbo coded MIMO OFDM system for receiver 1 and 2, with 2x2 transmit and receive antennas using BPSK modulation.

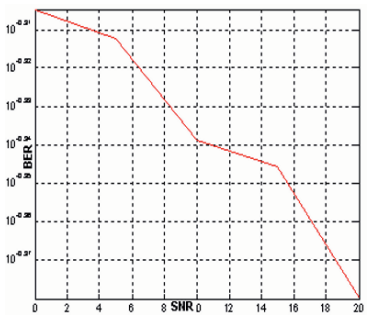


Figure 2:BER of turbo coded MIMO OFDM system with 2x2 Tx and Rx antennas using BPSK for receiver 1

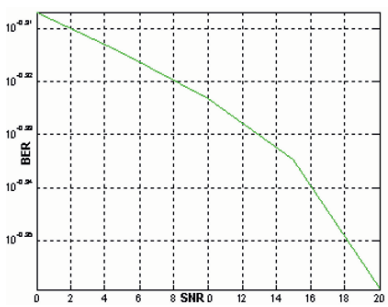


Figure 3:BER of turbo coded MIMO OFDM system with 2x2 Tx and Rx antennas using BPSK for receiver 2

We can observe from these two graphs that it attains BER of $10^{-0.37}$ at 18.5dB for receiver 1 and attain same BER at 16dB for receiver 2.

Figure 4, Figure 5 and Figure 6 shows BER of turbo coded MIMO OFDM system for receiver 1, 2 and 3, with 3x3 transmit and receive antennas using QPSK modulation.

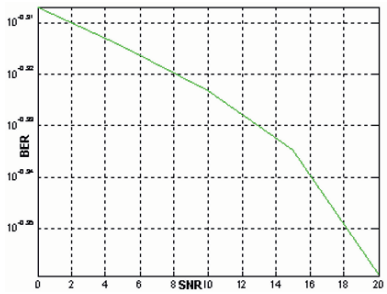


Figure 4:BER of turbo coded MIMO OFDM system with 3x3 Tx and Rx antennas using BPSK for receiver 1

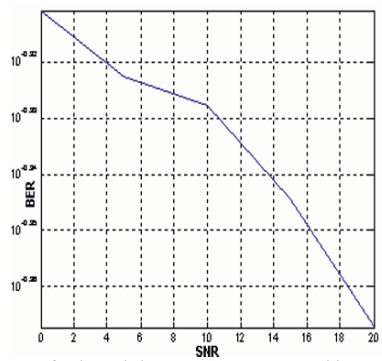


Figure 5:BER of turbo coded MIMO OFDM system with 3x3 Tx and Rx antennas using BPSK for receiver 2

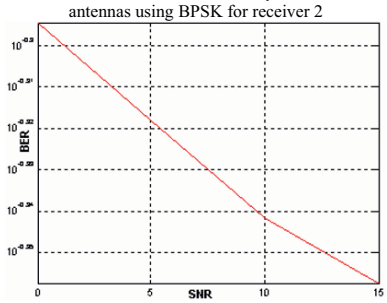


Figure 6:BER of turbo coded MIMO OFDM system with 3x3 Tx and Rx antennas using BPSK for receiver 3

We can observe from these three graphs that it attains BER of $10^{-0.35}$ at 18dB for receiver 1 and achieve same BER at 16dB and 12.5dB for receiver 2 and 3 respectively.

Figure 7 and Figure 8 shows BER of turbo coded MIMO OFDM system for receiver 1 and 2, with 2x2 transmit and receive antennas using QPSK modulation.

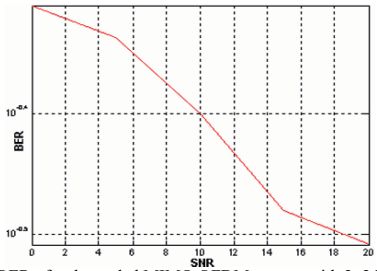


Figure 7:BER of turbo coded MIMO OFDM system with 2x2 Tx and Rx antennas using QPSK for receiver 1

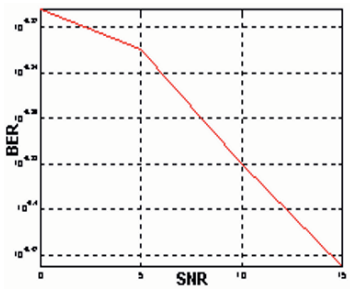


Figure 8:BER of turbo coded MIMO OFDM system with 2x2 Tx and Rx antennas using QPSK for receiver 2

We can observe from these two graphs that it attains BER of $10^{-0.5}$ at 18.3dB for receiver 1 and attain same BER at 14dB for receiver 2.

Figure 9, Figure 10 and Figure 11 shows BER of turbo coded MIMO OFDM system for receiver 1 and 2, with 3x3 transmit and receive antennas using QPSK modulation.

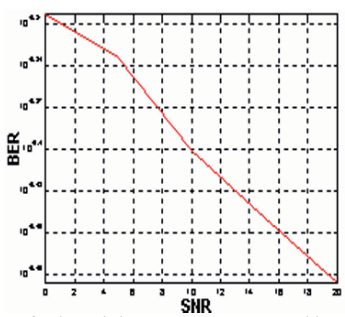


Figure 9:BER of turbo coded MIMO OFDM system with 3x3 Tx and Rx antennas using QPSK for receiver 1

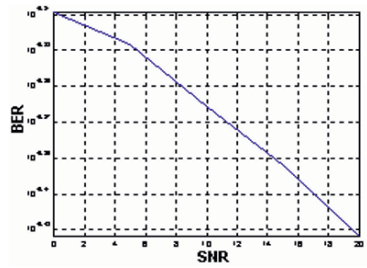


Figure 10:BER of turbo coded MIMO OFDM system with 3x3 Tx and Rx antennas using QPSK for receiver 2

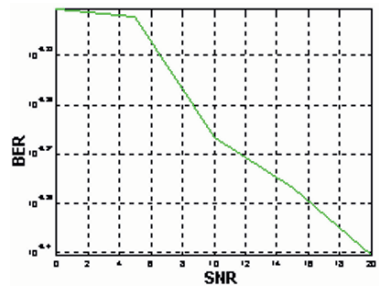


Figure 11:BER of turbo coded MIMO OFDM system with 3x3 Tx and Rx antennas using QPSK for receiver 3

We examine from these three graphs that it attains BER of $10^{-0.41}$ at 17dB for receiver 1 and achieve same BER at 20dB for receiver 2 and achieve BER $10^{-0.2}$ for receiver 3 at approximately 20dB.Figure 12 and Figure 13 shows BER of turbo coded MIMO OFDM system for receiver 1 and 2, with 2x2 transmit and receive antennas using 8PSK modulation.

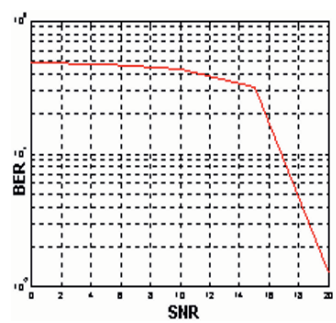


Figure 12:BER of turbo coded MIMO OFDM system with 2x2 Tx and Rx antennas using 8PSK for receiver 1

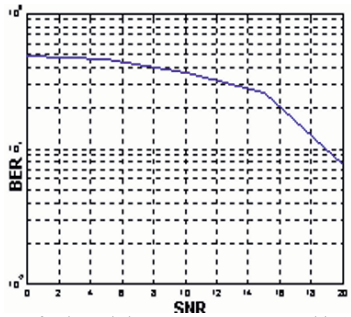


Figure 13:BER of turbo coded MIMO OFDM system with 2x2 Tx and Rx antennas using 8PSK for receiver 2

We can observe from these two graphs that it attains BER of $10^{-0.1}$ at 19dB for receiver 1 and attain same BER at 17dB for receiver 2.

Figure 14,Figure 15 and Figure 16 shows BER of turbo coded MIMO OFDM system for receiver 1, 2 and 3, with 3x3 transmit and receive antennas using 8PSK modulation.

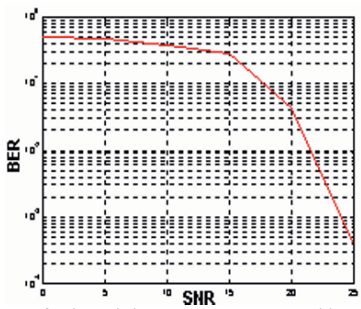


Figure 14:BER of turbo coded MIMO OFDM system with 3x3 Tx and Rx antennas using 8PSK for receiver 1

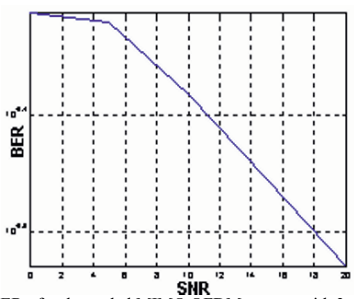


Figure 15:BER of turbo coded MIMO OFDM system with 3x3 Tx and Rx antennas using 8PSK for receiver 2

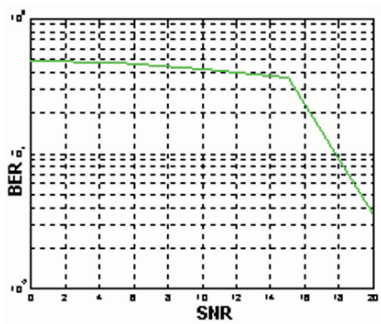


Figure 16:BER of turbo coded MIMO OFDM system with 3x3 Tx and Rx antennas using 8PSK for receiver 3

We examine from these three graphs that at 18 dB SNR its BER is $10^{-0.4}$ for receiver 1 and for receiver 2 at the same SNR it achieves BER of $10^{-0.5}$ and attain BER of $10^{-0.1}$ for receiver 3 at same SNR.

V CONCLUSION

This paper evaluates the performance of a working MIMO OFDM system with turbo codes. Analysis of simulation results shows that as we increase the number of transmit antennas the BER decrease for a given SNR value. It is observed that there is a considerable improvement in the BER when we increase the modulation.

REFERENCES

[1] Lim Wei Chee, B Kannan, Francois Chin, "MIMO Capacity Performance for Both Narrowband and Wideband Systems", 8th international conference on communication system, vol.1,pp. 426-430, Nov. 2002.

[2] M. Karkhanechi, Bemard Levy, "Spatial Multiplexing and Diversity Gain in OFDM-Based MIMO Systems", IEEE topical conference on wireless communication technology, pp. 299-301, Oct. 2003.

[3] Y. G. Li, J. H. Winters, N. R. Sollenberger, "MIMO-OFDM for wireless communications signal detection with enhanced channel estimation," IEEE transactions on communication, vol. 50, pp. 1471-1477, Sept. 2002.

[4] Maleki-Tehrani, A Hassibi, Cioffi, J.M, "Adaptive equalization of multiple-input multiple output (MIMO) frequency selective channels", conference on Signals, Systems, and Computers, vol. 1, pp. 547 –551, 1999.

[5]Simon Haykin, "Adaptive Filter Theory", Pearson Education, 3rd edition, ISBN: 81 7808 565 8, pp.339-400, 2003

[6] S.LGoff, Glavieux, Berrou, "Turbo-codes and high spectral efficiency modulation", IEEE international conference on communication, vol. 2, pp 645-649, May 1994.

[7] U.Wachsmann and J.Huber, "Power and bandwidth efficient digital communication using turbo codes in multilevel codes", European transaction on telecommunications, vol. 6, pp. 557-567, 1995.

[8] Pavel Loskot Norman C. Beaulieu, "Performance Analysis of Coded MIMO-OFDM Systems Over Generalized Ricean Fading Channels", Conference on Electrical and Computer Engineering, pp. 1634-1639, 2006

Dynamic Rate Control Algorithm for Streaming Media Over Wireless Channel

Kostas. E. Psannis

Abstract—**This paper presents a very efficient approach for streaming wireless video. The method is based on encoding MPEG video streams using differently encoding technique, I-P(Intra)-B(Intra). P(Intra) frames are coded using motion estimation and each one has a dependency only on to the preceding I- frame. Similarly, B(Intra) frames are coded based on a forward prediction from a previous I – frame as well as a backward prediction from a succeeding I – frame of the next Group Of Pictures (GOP). The requirements of wireless video include controlled jitter, improved packet loss characteristics and mechanism to dynamically adjust the output bit rate of the video encoder to meet the variable wireless channel throughput. Mechanisms for controlling the proposed MPEG-based streams using dynamic rate control algorithm are also presented and their effectiveness is assessed through extensive simulations. Compared with the MPEG-4 rate control algorithm, our proposed algorithm can effectively enhance loss characteristics, adjust the out video bit rate to meet the variable channel throughput channel; thereby, improve the smoothness of the video.**

Index Terms—**Wireless Video, Packet Loss Characteristics, Control Jitter, Smoothness of the Video**

I. Introduction

Streaming is the ability to commence processing data before all of it is available, thus enabling its delivery in real time. A video streaming system presents several challenges, such as the requirements for high video storage and high network bandwidth [1]-[3]. The properties of streaming media include the following. First the media is encoded without the knowledge of run-time channel status during transmission. For this reason, the encoding is desired to be flexible, e.g. using scalability, in order to achieve graceful degradation during channel fluctuations. Second media steaming media application typically have pre-roll buffer, e.g., 5-20 s, allowing for rate-distortion optimization and data packet scheduling to adaptively select data to be transmitted, as the channel varies [4], [5]. Second media steaming media application typically have pre-roll buffer, e.g., 5-20 s, allowing for rate-distortion optimization and data packet scheduling to adaptively select data to be transmitted, as the channel varies [6]. Video communications is fundamentally different from data communication, since they are delay and loss sensitive.

Kostas. E. Psannis is with the Department of Technology Management, University of Macedonia, Greece, (e-mails: kpsannis@uom.gr, mobility2net@gmail.com).

Unlike data packets, late arriving video packets are useless to the video decoder. Therefore the retransmission techniques are not generally applicable to video communication applications with low delay requirements. There are additional challenges for supporting video communications especially over wireless networks. Due to the mobility nodes, the topology of the network may change frequently. Thus the established connection routes between senders and receivers are likely to be broken during video transmission, causing interruptions, freezes, or jerkiness in the received video signal. These constraints and challenges, in combination with the delay and loss sensitive of media streaming applications, make video communications a challenging proposition [7], [8].

Rate control is an important issue in both wired and wireless streaming applications. However, rate control schemes recommended by current compression standards such as MPEG-4 [9] and H.263 [10], are optimized for constant bit-rate channels, and the can not adapt themselves in time to the variation of the channel bandwidth. In addition a rate control algorithm has to jointly decide encoding parameters and estimate current conditions in order to optimize and estimate current channel condition in order to optimize its encoding performance. MPEG-based codec also provides tools for 'scalable' coding where useful video can be reconstructed from pieces of the total bitstream. The total bitstream may be structured in layers, starting with a base layer (that can be decoded by itself) and adding refinement layers to reduce quantization distortion or improve resolution. Various scalable video coding techniques have been developed rapidly in the past decade. Spatial and temporal scalable video coding techniques that provide video at different resolutions and frames rate, were accepted in some main video coding standards such as MPEG-4 and H.263++ [9]-[11].

In recent years several techniques for streaming wireless video have been devised [12]-[22]. A TCP Friendly Rate Control which is equation based over wired networks have been proposed in [12]-[13]. Several research techniques have been reported to improve the performance of TCP Friendly rate control over wireless networks. These methods either hide end-hosts from packet loss cased by wireless channel error, or provide end hosts the ability to distinguish between packet cased by congestion, and that caused by wireless channel error [14]. One other approach that has been shown to improve wireless video quality by efficiently utilizing transmission energy is joint source coding and transmission power to different regions of a sequence based on their relative importance. For example, less power may be used to transmit packets in a static background region in order to

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 409–414.

allocate more power to regions of the sequence that are more difficult to conceal. Several research works on the region based rate control have been reported [15]-[19]. The Lagrange multiplier rate approach is employed for rate control in region-based coding [15]. Although the complexity of Lagrange multiplier was reduces it is still major concern in real time wireless video. Other methods [16]-[18] adopted a heuristic scheme to decide the quantization parameters for different region in a frame. These approaches lack a quantitative method to perform bit allocation among different regions. This may cause improper quantization parameters which imposes unreasonable bits used for different regions. Another region-based approach [19] adopts the most effective criteria as quantitative factors to directly control bit allocations among different regions and employs the priority concept to further adjust bit allocation under different channel conditions. Moreover recent research shows that video traffic under time varying channels may be modeled as multifractal cascade accounting for the multiscaling property of the traffic [20], [21]. A rate- distortion optimized packet scheduling algorithm has been proposed in [22], for streaming media by generating a number of nested sub streams, with more important streams embedding less important ones in a progressive manner.

In this paper we propose a very efficient approach for supporting wireless video streaming. The proposed method is based on encoding the video streams using the I-P(I)-B(I) GOP pattern. The corresponding version is obtained by encoding P(I)- frames using motion estimation and each one has a dependency only on to the preceding I- frame . B(I)-frames are coded based on a forward prediction from a previous I – frame as well as a backward prediction from a succeeding I – frame of the next Group Of Pictures (GOP). The proposed encoding pattern confirms with the goal of improved packet loss characteristics; adjust the output bit rate of the video encoder to meet the variable wireless channel. Consequently improve the smoothness of the video.

We analyze the effects of wireless video streaming regarding the packet loss and synchronization delay impose by the variable wireless channel conditions. Mechanisms for controlling the video streams are also investigated.

The paper is organized as follows. In Section II the preprocessing steps required to support efficient wireless video streaming is detailed. In addition mechanisms for controlling the proposed video streams are also investigated to assess the overhead of packet loss characteristics and variable channel throughput. Section III includes the simulation results demonstrating the performance of the algorithm. Section IV concludes the paper with final observations.

II. Proposed Approach

In order minimize the synchronization waiting time and the number of discarded frames due to I- or P- or B- frames loss we propose the use of differently encoded video sequence.

The corresponding version is obtained by encoding the original uncompressed video data as a sequence of I- P(I)-B(I) frames. Effectively this results that loss of P- frame does not affect the next frame to be decoded. P(I) frames are coded using motion estimation and each one has a dependency on to the preceding I- frame. Similarly, each B- frame has a dependency on to the preceding I- frame and the following I-frame.

On the other hand, this encoding pattern increases the size of P(I) and B(I) – frames respectively in the GOP. Note that the increase of the size of P(I) and B(I) frames depends on the type of motion of the video trace (high motion, little motion, complex motion)

A P(I) -frame is not a full frame. Instead, it is a Predictive video frame. This means that the P-frame follows an I –frame and only stores the data that has changed from the preceding I-frame

The B(I) frames however, are coded based on a forward prediction from a previous I – frame as well as a backward prediction from a succeeding I – frame of the next GOP.

There is no defined limit to the number of consecutive B frames that may be used in a Group Of Pictures (GOP), and of course the optimal number is application dependent. Most broadcast quality applications however, have tended to use 2 consecutive B frames (M=3) as the ideal trade-off between compression efficiency and video quality. Figure 1 depicts the encoding and transmission order of the proposed encoding pattern (I-P(I)-B(I)).

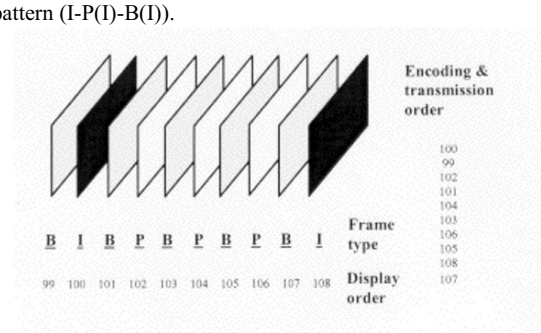


Fig. 1 Encoding and Transmission order N=9, M=3 , I-P(I)-B(I) structure

Current compression standards such as MPEG-4 and H.263, are optimized for constant bit-rate channels, and the can not adapt themselves in time to the variation of the channel bandwidth. In addition a rate control algorithm has to jointly decide encoding parameters and estimate current conditions in order to optimize and estimate current channel condition in order to optimize its encoding performance.

A common approach to control the size of an MPEG frame is to vary the quantization factor on a per-frame basis [23]. The quantized coefficients $QF[u,v]$ are computed from the DCT coefficients $F[u,v]$, the quantization_scale, $MQUANT$, and a quantization_matrix, $W[u,v]$, according to the following equation.

$$QF[u,v] = \frac{16 \times F[u,v]}{MQUANT \times W[u,v]}$$

The normalized quantization factor $w[u,v]$ is

$$w[u,v] = \frac{MQUANT \times W[u,v]}{16}$$

The quantization step makes many of the values in the coefficient matrix zero, and it makes the rest smaller. The result is a significant reduction in the number of coded bits with no visually apparent difference between the decoded output and the original source data. The quantization factor may be varied in two ways.

- Varying the quantization scale ($MQUANT$)
- Varying the quantization matrix ($W[u,v]$)

Our goal is to properly select the quantization factor on a per -P(I) and -B(I) frame basis in order to optimize its encoding performance according to variable channel condition and provide acceptable visual quality. For these reasons the encoding algorithm fulfils the following equation for the P(I) and B(I) frames respectively.

$$P(I)_{bits\_Size} \le Threshold^{P} \times I_{bits\_Size} \quad (1)$$

$$B(I)_{bits\_Size} \le Threshold^{B} \times I_{bits\_Size} \quad (2)$$

$$subject\_to$$

$$Thersholds^{P,B} = F(type\_of\_motion)$$

$$Thershold^{P,B} = F(Channel\_Throughput)$$

For the sake of simplicity, assume that the average video rate per GOP and the available channel throughput (C bits / sec) is equal. Thus the available channel throughput can be computed as follows

$$C = I_{bit\_Size} + (\frac{N}{M} - 1)P_{bit\_Size}(I) + \_\_\_\_\_$$

$$\_\_\_\_\_ + (N - \frac{N}{M})B_{Bit\_Size}(I)\ bits/sec$$

Note that the values of the $Thresholds^{P,B}$ depend on the type of motion of the original video (uncompressed) sequence (little motion, high motion, complex motion) and the available channel throughput. Little motion sequences means that more bits should be allocated to I- frames and less bits to the remaining P(I) and B(I) frames. On the other hand high motion video sequences means that P(I) and B(I) frames have more bits than the same frames in the little motion. Complex motion video sequence is the combination of the above on a per GOP basis. In order to achieve video streaming applications with available channel throughput and acceptable visual quality, computer simulation can be used in order to determine the values of the two thresholds ($Thresholds^{P,B}$).

Figure 2 depicts the selected $Thresholds^{P,B}$ for different types of motion and available channel throughput. We use Microsoft MPEG-4 Visual Reference software to encode video traces. All test sequences are in QCIF format (176 x 144 pixels/frame) and encoded at target frame rate 30 frames/s, the number of frames to be encoded is 150 frames. The first frame is intra frame and the remaining frames are P(I)- and B(I). The GOP Length is N=15 and distance between consecutive I- or P(I)-frames is M=3. We used three video traces with different type of motion (Talk Show- little motion, Animation-high motion, Music Video Clip-Complex Motion).

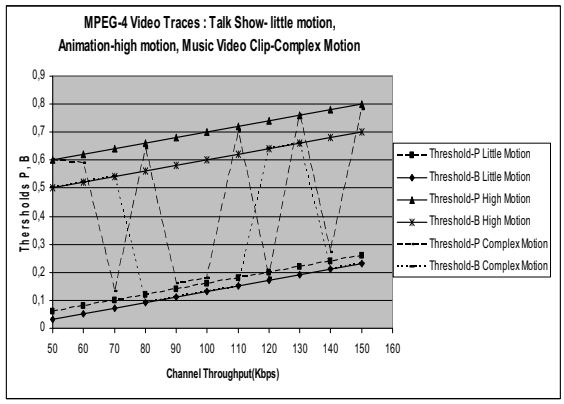


Fig. 2: Thresholds (P or B) as a function of Channel Throughput for different type of video traces

Figure 2 depicts the selected Thresholds for the P(I)- and B(I)-frames respectively as a function of the available channel throughputs and for different types of video traces.

In the first encoding attempt for the P(I) and B(I) respectively , the encoder starts with the nominal quantization value that was used to encode the previous I-frame. If the resulting frame size fulfils the (1) or (2) the encoder proceeds to the next frame P(I)- or B(I)- frame. Otherwise, the quantization factor (quantization_matrix, $W[u,v]$) varies and the same frame is re-encoded. The quantization matrix can be modified by maintaining the same value at the near-dc coefficients but with different slope towards the higher frequency coefficients. This procedure is repeated until the size of the compressed frame fulfils (1) for the P(I)- frame and (2) for B(I) frame respectively . The advantage of this scheme is that it tries to produce video streaming services with the same constant quality of I-P-B structure but when it is not possible it minimizes the fluctuation in video quality.

Fig. 3 shows the quantization_matrix $W[u,v]$ with the same value at the near-dc coefficients but with different slope towards the higher frequency coefficients. In other words, the quantization_scale is fixed $MQUANT$ =8, which is the same for the the I-frames and the quantization_matrix $W[u,v]$ varies.

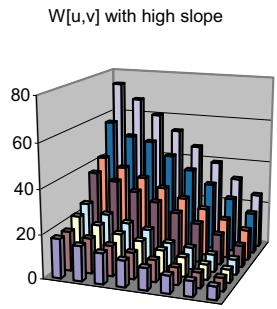


Fig.3: Normalized quantization matrices w[u,v] both with MQUANT=8 and W[u,v] with high slope.

III. EXPERIMENTS

There are two types of criteria that can be used for the evaluation of video quality; subjective and objective. It is difficult to do subjective rating because it is not mathematically repeatable. For this reason we measure the visual quality of the proposed encoding algorithm using the Peak Signal-to-Noise Ratio (PSNR).

All test sequences are in QCIF format (176 x 144 pixels/frame) and encoded at target frame rate 30 frames/s, the number of frames to be encoded/decoded is 150 frames. The first frame is intra frame and the remaining frames are P(I)- and B(I). The GOP Length is N=15 and distance between consecutive I- or P(I)-frames is M=3. We used three video traces with different type of motion (Talk Show- little motion, Animation-high motion, Music Video Clip-Complex Motion) at 100 Kbps channel throughput. The PSNR is obtained by comparing the original raw frame with its decoded version with encoding being done using the proposed encoding algorithm.

Figure 4 depicts the PSNR for the encoded frames using the proposed encoding algorithm for different types of video sequences. The average PSNR value for the 150- frames is 39,8dB for the little motion video sequence, 38.86 dB for the high motion video sequence and 38,875 for the complex motion video sequence. The average quality is better under the little motion video sequence.

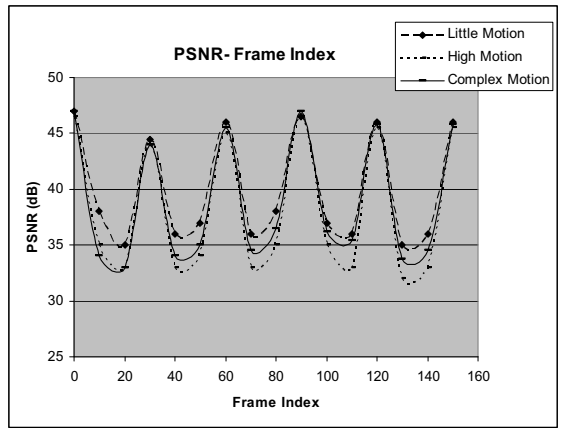


Fig. 4: PSNR as a function of frame index for different types of video sequences

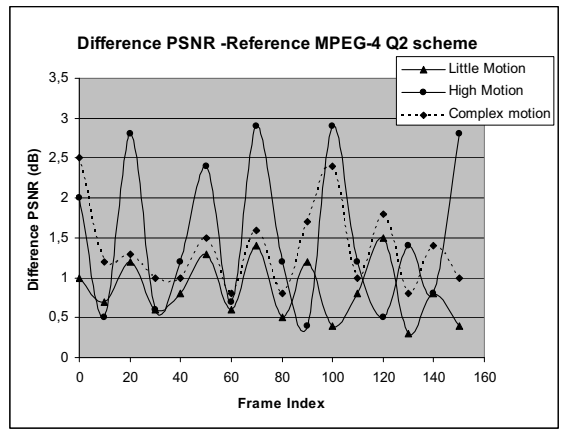


Fig. 5: Difference PSNR as a function of frame index for different types of video sequences

The absolute values of PSNR do not convey the advantage of the proposed encoding algorithm. For this purpose, we compute the PSNR values for the 150 frames when encoding is done with MPEG-4 Q2 scheme [24] and uses these values as a reference. For each frame, we compute the difference between its reference PSNR value and the PSNR value resulting using the proposed encoding algorithm for the different types of video traces. . These differences are plotted in Fig. 5. In the same figure a large value indicates a large deviation from the reference PSNR, and thus lower quality. Clearly the proposed encoding algorithm achieves better visual quality for the little motion video sequence due to the less bit allocation of the P (I) and B(I) frames respectively

Figure 6 depicts the performance of the proposed encoding algorithm compared with the performance MPEG-4 Q2 scheme [24] codec during frame loss. The percentage of

frames that are dropped is varied and it is clearly seen that the proposed encoding algorithm maintains the visual quality for the different types of video sequences. It should be emphasized that the MPEG-4 Q2 scheme codec, the quality degrades with the increase in frame loss percentage.

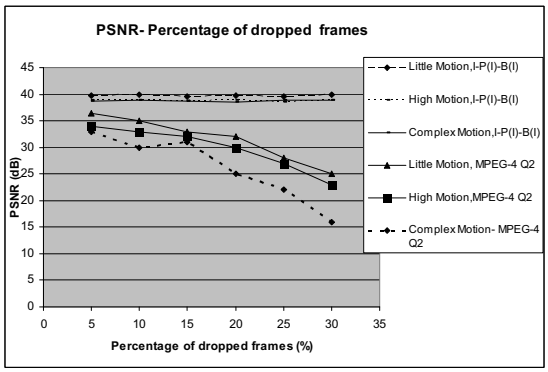


Fig. 6:PSNR as a function of dropped frames for the proposed encoding algorithm and the reference MPEG-4 Q2 scheme

IV. CONCLUSIONS

The requirements of wireless video include controlled jitter, improved packet loss characteristics and mechanism to dynamically adjust the output bit rate of the video encoder to meet the variable wireless channel throughput. In this paper, we investigated the constraints of supporting video streaming, regarding the delay and loss sensitive imposes by wireless time varying channels. In order to overcome these constraints we proposed the use of differently encoded version of each video sequence. Each one of the differently coded sequences is obtained by encoding the using I-P(I)-B(I) GOP pattern. Mechanisms for controlling the proposed video streams using dynamic rate control algorithm are also presented. Experimental results shows that the proposed method achieves acceptable visual quality compared to MPEG-4 Q2 scheme and improves packet loss characteristics for different types of video traces; thereby, improve the smoothness of the video. Future work involves optimization over combined networks from wireline to wireless links, optimization with VCR-Full interactive functionality [23].

REFERENCES

[1] Z. He, J . Cai and C. W. Chen , “Joint source channel rate-distortion analysis fro adaptive mode selection and rate control in wireless video coding “, IEEE Trans. Circuits Syst. Video Technology, vol 12, no. 6, pp 511-523, Jun 1999.

[2] M. Etoh and T. Yoshimura, “Advances in wireless video delivery,” Proceedings of the IEEE, vol. 93, no. 1, pp. 111–122, 2005.

[3] D. -P. Wu, Y. -T. Hou, W. Zhu, Y.-Q. Zhang, and J. Peha, "Streaming Video over the Internet: Approaches and Directions," IEEE Trans on Circuits and Systems for Video Technology, Feb 2001, pp 1-20.

[4] G. J. Conklin, G. S. Greenbaum, K. O. Lillevold, A. F. Lippman, and Y. A. Reznik, “Video Coding for streaming media delivery on the Internet, “, IEEE Trans. Circuits Syst. Video Technol., Mar 2001, pp 269-281.

[5] N. Cranley, L. Fiard and L. Murphy, "Quality of Service for Streamed Multimedia over the Internet", Proc. Irish Signals and Systems Conference 2000, Dublin, Ireland, June 2000

[6] M. Chen and A. Zakhor, "Rate Control for Streaming Video over Wireless", IEEE Wireless Communications, Vol. 12, Issue 4, Aug. 2005, pp. 32-41

[7] T. Stockhammer and M. M. Hannuksela, “H.264/AVC video for wireless transmission,” IEEE Wireless Communications, vol. 12, no. 4, pp. 6–13, 2005

[8] G. De Los Reyes, A. R. Reibman, S.-F. Chang, and J. C.-I. Chuang, “Error-resilient transcoding for video over wireless channels,” IEEE Journal on Selected Areas in Communications, vol. 18, no. 6, pp. 1063–1074, 2000

[9] Coding of Moving Pictures and Associated Audio.MPEG98/W21994, (MPEG-4), Mar.1998.

[10] ITU-T, Recommendation H.263 : Video Video Coding for Low Bit Rate Communications, version 2, March 1993

[11] MPEG video group. Information Technology, Generic coding of moving pictures and associated audio, ISO/IEC 13818-2, International standard, 1995

[12] S. Floyd, M. Handley, J. Padhye, and J Widmer, “ Equation-based Congestion control for unicast applications, “ in Proc. ACM SIGCOMM, Stockholm, Sweden, Aug. 2000, pp. 43-56

[13] S. Floyd and K.Fall, “ Promoting the use of end-to-end congestion control in the internet”, IEEE/ACM Trans. Networking, no.4, pp. 458-472, Aug. 1999.

[14] M. Chen and A. Zakhor, "Multiple TFRC Connections Based Rate Control for Wireless Networks," IEEE Trans. on Multimedia, Vol. 8, No. 5, Oct. 2006, pp. 1045-1062

[15] H. Song and C.-C. Jay Kuo, “A region-based H.263+ codec and its rate control for low VBR video,” IEEE Trans. Multimedia, vol. 6, no. 3, pp. 489–500, Jun. 2004.

[16] C.-H. Lin and J.-L. Wu, “Content-based rate control scheme for very low bit-rate video coding,” IEEE Trans. Consumer Electron., vol. 43, no. 2, pp. 123–133, May 1997.

[17] S. Aramvith, H. Kortrakulkij, D. Tancharoen, and S. Jitapankul, “Joing source-channel coding using simplified block-based segmentation and content-based rate-control for wireless video transport,” in Proc. Int. Conf. Information Technology: Coding and Computing (ITCC) 2002, Las Vegas, Apr. 2002, pp. 71–76.

[18] C.-W. Lin, Y.-J. Chang, and Y.-C. Chen, “A low-complexity face-assisted coding scheme for low-bit-rate video telephony,” IEICE Trans. Inform. Syst., vol. E86-D, no. 1, pp. 101–108, Jan. 2003.

[19] Yu Sun, Ishfaq Ahmad, Dongdong Li, , and Ya-Qin Zhang, , “Region-Based Rate Control and Bit Allocation for Wireless Video Transmission”, IEEE Trans on Multimedia, vol. 8, no.1, Feb 2006, pp 1-10 .

[20] C. A. V. Melo and N. L. S. Fonseca, “An envelope process for multi-fractal traffic modeling,” in Proc. IEEE Int Conf. Communication (ICC) 2004, vol. 4, June 20–24, 2004, pp. 2168–2173.

[21] A. Erramilli, O. Narayan, A. Neidhardt, and I. Saniee, “Multi-scaling models of TCP/IP and sub-frame VBR video traffic,” J. Commun. Networks, vol. 3, no. 4, pp. 383–395, Dec. 2001.

[22] S. H. Kang and A. Zakhor, "Effective Bandwidth Based Scheduling for Streaming Multimedia", IEEE Transactions on Multimedia, Vol. 7, No. 6, December 2005, pp. 1139-1148

[23] K. E. Psannis, M. G. Hadjinicolaou, and A. Krikelis, “MPEG-2 streaming of full interactive content,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 2, pp. 280–285, 2006.

[24] H.-J. Lee, T. Chiang, and Y.-Q. Zhang, “Scalable rate control for MPEG-4 video,” IEEE Trans. Circuits Syst. Video Technol., vol. 10, pp. 878–894, Sep. 2000.

APPENDIX

In the appendix, some P(I) and B(I) frames are depicted for the different types of motion video traces (Talk Show- little motion, Animation-high motion, Music Video Clip-Complex Motion)

(a) 33rd P(I)- frame

(b) 61rd B(I)-frame

Figure 7. Subjective results of the 33rd P(I)- frame (a) and the 61rd B(I)-frame for the Talk-Show video trace (Little Motion)

(a) 33rd P(I)- frame

(b) 61rd B(I)-frame

Figure 8. Subjective results of the 33rd P(I)- frame (a) and the 61rd B(I)-frame for the Animation video trace (high motion)

(a) 33rd P(I)- frame

(b) 61rd B(I)-frame

Figure 9. Subjective results of the 33rd P(I)- frame (a) and the 61rd B(I)-frame for the Music Video Clip (Complex Motion)

Interactive Compression Algorithms for Streaming Media Over High Speed Networks

Kostas. E. Psannis

***Abstract*—This paper presents interactive compression algorithms for streaming media over high speed networks. An MPEG coded media sequence is typically partitioned into small intervals called GoP (Group Of Pictures). This structure allows a simple realization of forward (normal)-pay operation but imposes several additional constraints on the other interactive functions. Implementation of the rewind operations requires that the decoder either decodes the whole GoP and store it, or it decodes the GoP up to the current frame to be displayed. Both of these techniques lead either to the requirement for large storage at the client machine (to store a fully decoded GoP) or massive decoding processing power (to fully decode a GoP at required displayed frame rate). Neither of these options is desirable. Furthermore, Fast Play (Fast Forward/Jump Forward) functions of MPEG coded video present are problematic. When a P-/B-frame is requested, all the related previous P-/I-frames need to be sent over the network. This is likely to lead in considerable increase in communication bandwidth and decoding power compared to the normal-play mode. Several interactive algorithms are detailed with respect to server load /network bandwidth and decoder complexity. Finally an extensive comparative study has been carried out in order to determine the trade offs of the proposed approaches.**

***Index Terms*—Compression Algorithms, Interactive Video, VCR functionality, Streaming Interactive Media**

I. Introduction

Today's media technology herald an exciting era that will enormously impact daily life. Media is any combination of text, graphics, audio, video, animation and data.. Moreover Digital Video Cassette Recording (VCR) functionality enables quick and user-friendly browsing of media content, thus making it a highly desirable feature in streaming video applications. Interactive access to video content is generally defined as a program or service controlled by the user and which can affect the content itself, the presentation manner of the content, or the presentation order of the content. Full range of interactive functions include play/resume, stop, pause, Jump Forward (JF)/ Jump Backward (JB), Fast Forward (FF)/ Fast Rewind FR), Slow down (SD), and Slow Reverse (SR), Rewind [1]-[5]. The difficulty of supporting interactivity varies form one interactive function to another. A stop or pause followed by resume is relatively easy to support since there is no requirement for more bandwidth than is already allocated for normal playback.

Kostas. E. Psannis is with the Department of Technology Management, University of Macedonia, Greece, (e-mails: kpsannis@uom.gr, mobility2net@gmail.com).

However, Fast Scanning {(FF), (FR), (JF), (JB)) functions involve displaying frames at multiple times the normal rate. Transporting and decoding frames at such high-rate is prohibitively expensive and is not feasible using the hardware decoders available today or in the foreseeable future[5]-[7]. Implementation of the full interactive functions with the MPEG coded video presents a number of considerable challenges associated with video data storage and playout. An MPEG video stream comprises intra-frames (I), predicted frames (P), and interpolated frames (B). According to MPEG coding standards, I-frames are coded such that they are independent of any other frames in the sequence; P-frames are coded using motion estimation and each one has a dependency on the preceding I- or P- frame; finally the coding of B-frames depends on the two "anchor" frames - the preceding I/P frame and the following I/P frame [6]-[8].

In recent years several techniques for supporting interactivity for MPEG code video streaming applications have been devised [9]-[18]. In [9], [10] and {11] interactive functions are supported by dropping parts of the original MPEG-2 video stream. Typically, this is performed once the video sequence is compressed and aims to reduce the transport and decoding requirements. These approaches introduce visual discontinuities during the interactive mode, due to the missing video information. Alternatively interactive functions can also be supported using separate copies of the video streams that are encoded at lower quality of the normal playback copy [12], [13]. In these cases, there is no significant degradation in the visual quality. However, the number of pre-stored copies of the movie limits the speed-up granularity. Moreover, [14] proposed a differently encoding pattern of the media traces. Full interactive functions are produced by encoding every N^{th} frame of the original media traces as a sequence of I-P(M) frames. This method achieves minimum additional resources at the server load/network bandwidth\decoder complexity and acceptable visual quality at the client's end. In addition this method [14] supports full range of interactive functions.

The paper is organized as follows. In Section II proposed algorithm that support interactive functionality of the media streams are detailed. In Section III a comparative study has been carried out in order to determine the tradeoffs of the proposed algorithms with respect to additional resources (server load/ network bandwidth /decoder complexity), range of interactivity and visual quality. Section IV concludes the paper with final observations.

II. Proposed Algorithms

There are three different rates at which a video can flow

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 415–420.

from the disk of the server to the client display which should be taken into consideration while analyzing interactive functions. First, the video data has a retrieval rate going from the server's disk array to the main memory. Second, the data has a network transmission rate going from the server's memory to the client's memory. Lastly, the client decoder has a consumption rate at which it displays data from the client's memory. This section discusses and analyses the various proposed schemes for supporting interactive operations, according to the above requirements in the MPEG-based media encoding

A. Drop parts of the original video stream [9]-[11]

Interactive operations can be supported by dropping parts of the original video stream. Dropping aims to reduce both the transport and decoding requirements of the interactive mode without causing significant degradation in the video quality.

1) Send only I - P frames[9]

This method sends only the I- and P-frames to support Fast Forward operations. Transmitting only the I- and P- frames of each GoPs a limited number of speedups can be achieved. The main problem with this method is that it increases the load on the network (the frame rate remains constant).

For example consider a movie that is encoded into 3 GoPs, where the GoPs Length is N=9 and the distance between the I-I/P frames is M=3. The total number of I-P-B frames are in the 3 GoPs are, $I=3$, $P=6$, $B=18$. The authors in [9] propose a storage pattern in order to minimize the disk overhead occurred by accessing k-sequences every second. Since the data being displayed is not contiguous, the authors recommend that the MPEG-based file is reordered. If the data were kept contiguous on the disks, then a disk rotation, seek and transfer must be performed for each frame, which increases the disk overhead in order to support faster playback.

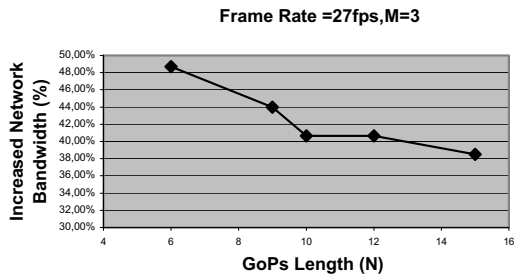


Figure 1: The increased percentage of the network bandwidth as a function of GOP Length (N).

The graph in figure 1 shows that there is not a linear increase (%) in the network bandwidth during the Fast Forward (FF) mode as a function of N. For Fast Reverse playback operation the authors in [9] propose to send only the I-frames in reverse order. P-frames are based on forward prediction only and cannot be shown backwards. When the user requests a fast reverse mode the server must reschedule the disk to the desired frames in reverse order. The main problem is that the server, for a speedup of k, needs to read all I-frames from the k- GoPs. In addition, sending only the I-frames results in the enormous increase in the network bandwidth

2) GOP-Skipping [10]

The GoP-Skipping method proposed in [10] maintains the same playback frame rate and skip entire GoPs.

(a) Normal playback

$$1-GoPs\ \ 2-GoPs\ \ 3-GoPs.....n-GoPs$$

(b) Fast Forward mode (k times the normal rate)

$$1-GoPs\ \ k-GoPs\ ....n-GoPs$$

To achieve speedup k-times the normal play, every k-times GoPs is sent and consumed by the client's decoder. This method can achieve variable speedups. Figure 2 shows the variable number of speedups as a function of skipped GoPs for N=15 and frame rate 30 fps.

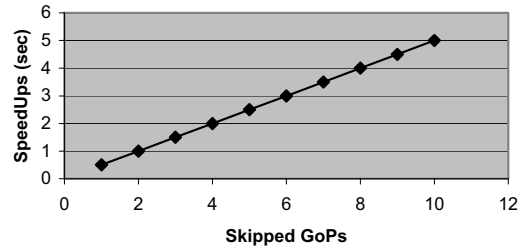


Figure 2: Number of speedups as a function of skipped GoPs

One limitation of this approach is that only independently decode-able GoPs can be used. If open GoPs are used, some extra processing must occur at the client-site increasing the requirements of the decoder complexity. Consider the display sequence of an open GoPs with GoPs length, N=6 and M=3.

(a) Display order

$$I_1B_2B_3P_4B_5B_6I_7B_8B_9P_{10}B_{11}B_{12}I_{13}B_{14}B_{15}P_{16}B_{17}B_{18}I_{19}.....$$

(b) Transmission order

$$I_1P_4B_2B_3I_7B_5B_6P_{10}B_8B_9I_{13}B_{11}B_{12}P_{16}B_{14}B_{15}I_{19}B_{17}B_{18}.....$$

For speedup k=2, two GoPs are skipped. In this case frames B_5 and B_6 would have to be discarded by the client's decoder since the forward I-frame I_7, that they depend on would be skipped. This phenomenon would result in slight hiccups at the client's set top box. It is worth mentioning that there are no increases in the resources at the server, network bandwidth and client's decoder when users switch to Fast Forward/ Fast Rewind. In the case of rewind, current MPEG-based hardware decoders cannot play frames of an MPEG-

based GoPs in reverse order. Thus, the GoPs are sent in reverse order but the individual frames are played in forward order. The visual output will not look faster since the playback rate of each GoPs does not change. This however, may be acceptable to the end user who is searching backward for a particular part of a video. To demonstrate another drawback of this method, let a GoPs consist of 15 frames, the normal playback rate be 30fps and the speedup rate be four. Thus from the original stream, 15 frames are displayed, 120 are skipped, 15 are displayed, etc. This is a drawback because if the video has complex motion within the four seconds, then the information will be lost. Hence this method cannot be recommended because the visual quality will be degraded.

3) Partial GoPs-skipping [11]

Another approach, proposed in [11], is to partial skip GoPs rather than skip full GoPs. Consider the display sequence of an open GoPs with GoPs length, N=15 and M=3.

(a) Normal playback

$I_1 P_4 B_2 B_3 P_7 B_5 B_6 P_{10} B_8 B_9 P_{13} B_{11} B_{12} I_{16} B_{14} B_{15} P_{19} B_{17} B_{18}$.....

(b) Fast playback

$I_1 P_4 B_2 B_3 I_{16} P_{19} B_{17} B_{18}$.....

The first four frames are considered as an independent sequence because they can be decoded independently. The first 7 or 10 or 13 frames, can be considered as an independent sequence. For fast playback, independent sequences are retrieved and sent while the rest of the frames in each GoP are skipped. When using this method there are only a fixed number of speedup rates than can be achieved. Figure 3 depicts the number of speedups as a function of transmitted frames.

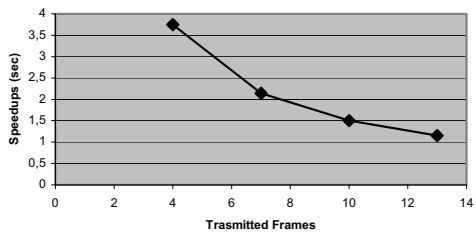


Figure 3: Relative increase in the speedups as a function of the transmitted frames

Like the skipping segment method, the same playback rate is maintained at the client. The bit rate required to be sent over the network will be increased, as well as the resources at the server.

Consider a movie that is encoded into 2 GoPs containing the following frames.

(1GoPs): $I_1 B_2 B_3 P_4 B_5 B_6 P_7 B_8 B_9 P_{10} B_{11} B_{12} P_{13} B_{14} B_{15}$

(2GoPs): $I_{16} B_{17} B_{18} P_{19} B_{20} B_{21} P_{22} B_{23} B_{24} P_{25} B_{26} B_{27} P_{28} B_{29} B_{30}$

Each GoPs is 15 frames long (N=15). Like the skipping method, this method maintains the same playback rate at the client's end. Thus, instead of sending 2 GoPs (30 frames/ sec) consisting of I- P - B frames during each cycle as in normal playback, the first 4 frames are delivered.

$I_1 B_2 B_3 P_4 \quad I_{26} B_{17} B_{18} P_{19} \quad I_{31} B_{31} B_{33} P_{34} \quad I_{46} B_{47} B_{48} P_{49}$

1 GoPs 2 GoPs 3 GoPs 4GoPs

The average size of the first four frames or the fist seven frames of each GoP would be:

$$Average(I\,BB\,P)Size = \frac{I_{average} + P_{aveage} + (2 \times B_{aveage})}{4}$$

$$Average(I\,BBPBBP)Size = \frac{I_{average} + (2 \times P_{aveage}) + (4 \times B_{aveage})}{7}$$

Figures 4 shows the increase in the bit rate as a function of the supported speedups for GoPs Length N=15 and M=3 using typical characteristic of MPEG-2 compressed video (Appendix B.3).

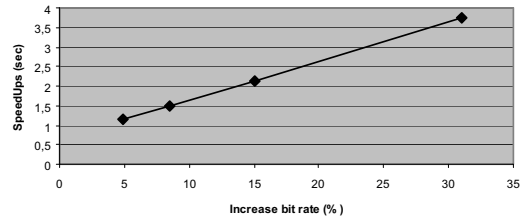


Figure 4: Relative increase in the bit rate as a function of the supported speedups

In addition to the higher bandwidth over the network required over the Fast Forward mode, extra overhead may be incurred by the I/O subsystem. If a GoP is used as the retrieval block, normally the GoP is stored contiguously on disk.

Assuming that during each I/O cycle 1 GoP (15 frames) is retrieved, the disk executes one rotation, seek, and transfer. If four different frames from three GoPs (12 frames) and three frames from the next GoPs are retrieved during each cycle, the disk must perform four rotations, seeks, and retrievals for every 15 frames retrieved. This procedure increases the overhead in processing the fast mode request. It is worth mentioning that using IBBP structure rewind functions can no be implemented due to the P-, and B- frames dependencies.

B. Separate Copies of the movie [12],[13]

Instead of dropping frames after compression, some researchers suggested supporting interactive operations using separately copies of the movie that are encoded at lower quality than the quality of the normal playback.

1) Skipping raw of frames [12]

This approach is based on encoding separate copies of the movie to be used for interactive operations in a Video On Demand (VOD) system. Each copy is generated by skipping rows of frames before compression. The server maintains

multiple different encoded versions of each movie. The normal version is used for normal-speed playback. The other versions, which are referred to as the scan version, are used for Fast Scanning (FS) operations. Each scan version is used to support both Fast Forward Scanning (FFS) and Backward Fast Scanning (BFS). For a given speedup factor, the corresponding scan version is obtained by encoding a subset of the raw uncompressed frames of the original movie at a sampling rate of 1-to-s, where s is the skip factor. The skip factor defines the sampling rate. Scan versions are encoded in a way such that, when played back at the normal frame rate, they give a perception of a faster video in the forward or backward direction. The server switches between various versions depending on the requested interactive operations (only one version is transmitted at a given instant time).

It is worth mentioning that the interactive scanning operations] can be supported with some extra network bandwidth. Encoding a sample of the raw frames may result in higher values of $P_{average(s)}, B_{average(s)}$. The encoder can control the size of P-B frames in the scan mode using two predefined thresholds.

$$T(P) = P_{\max}(1 + S_P)$$
$$T(B) = B_{\max}(1 + S_B)$$

where $P_{\max}$ and $B_{\max}$ are for normal version and S_P, S_B are nonnegative constants. A P (B) frame is encoded such that its size is no higher than $T(P)\{T(B)\}$. On the other hand this results in variable video quality during Fast Scanning operations. Generating separate copies for Fast Scanning operations comes at the expense of extra storage of the server and some variability in the quality of the motion picture during the Fast Scanning (FS) periods. Figure 5 depicts the percentage of the increase in the storage overhead as a function of N using typical characteristics of MPEG-2 compressed video with skip factor $s = 4$ and $S_P = S_B = 0$. For n scan versions (variable speedups) with skip factors $s_1\, s_2 \ldots\ldots\ldots\ldots s_n$, the relative increase in the storage requirement is given by $\dfrac{\sum_i W_{scan}(s)}{W_{normal}}$.

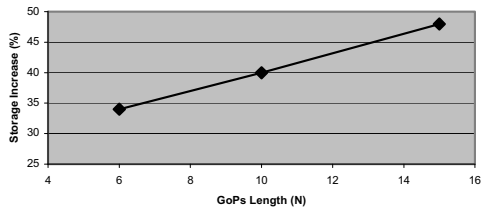


Figure 5: Relative increase of the storage overhead as a function of GoPs Length (N)

2) Alternative special file [13]

In this method a special file is created specifically for use in Fast Forward/ Fast Rewind mode and multi-resolution viewing. The encoding of the special file can be done in two ways. If the original uncompressed file is available, then every n-th frame is encoded for a speedup of n. If the original file does not exist, the compressed stream is first uncompressed and every nth frame is then re-compressed. The alternative file can be created based on the motion within the original file. During periods of little motion, large numbers of frames are skipped from the original file. During periods of high motion, lower numbers of frames are skipped.

The special file is created to be approximately 20-25% of the original file so that the average bit rate of the resulting file is less than or equal to the average bit rate of the original. This guarantees that, when switching form normal playback to fast playback, there is no increase in the load on the server or the network. The special file is encoded at both lower temporal resolution and spatial resolution than the original file.

In addition, when creating the special file the authors in [13] propose to update the bit-stream so that it can support both the Fast Forward (FF) and Fast Rewind (FR) mode. To solve the problems of dependencies in MPEG-2 frames the authors propose two strategies. First, macroblocks are forced to be either intra-coded (I) or purely interpolated (B). P macroblocks are not acceptable because they have a unidirectional dependency, which would make it impossible to play frames back in reverse directions. This however, would not entirely solve the problem since the interpolated B macroblocks have different distance vectors with respect to the two reference frames. Hence if the reference frames are presented in the reverse order (as in the case of rewind), the decoding would be incorrect. To overcome this problem, the distance vector are forced to be equal in both directions, and set equal to the smaller of the two vectors. This makes the interpolated frames symmetrical with respect to both the forward and backward reference frames and enables them to be decoded correctly even if the reference frames are presented in reverse order. Figure 6 depicts the increase in the storage at the server as a function of GoPs (N), using typical characteristics of MPEG-2 compressed video (Appendix B.3) with the frame format of the special file $I\,B\,B\,B\,B\,I$ and encoding every third frame of the original file.

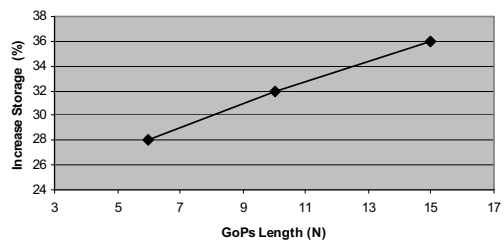


Figure 6: Relative increase of the storage as a function of N

The main advantages of using this method are that it does not increase the load on the server since it can be created with the same playback rate as the original file. Also the file can be created, as described above, such that it can be shown forward and backward and in multi-resolution viewing. The main disadvantage of using this scheme is that a separate file must be created for each speedup rate, increasing the amount of storage used in the system and hence an increase in cost.

C. I-P(Marionnete) Frames [14]

In this efficient approach the server maintains multiple differently encoded versions of each video streams in order to support interactive functionality. . One version, which is referred to as the normal version is used for normal-speed playback. The other versions are referred to as interactive versions. Each interactive version is used to support Fast/Jump Forward/Backward Slow Down/Reverse and Reverse at a variable speedup. The server switches between the various versions depending on the requested interactive function. Assume that I- frame is always the start point of interactive mode. Since I- frames are decoded independently, switching from normal play to interactive mode and vice versa can been done very efficiently. Note that only one version is transmitted at a given instant time.

The corresponding interactive version is obtained by encoding every N-th (i.e., uncompressed) frame of the original movie as a sequence of I- P(Marionette) - frames $(N_{\text{int}\,eractive} = \text{var}\,iable, M_{\text{int}\,eractive} = 1)$. Effectively this results in repeating the previous I-frame in the decoder, enhancing the visual quality during the interactive mode. This is because it freezes the content of the I-frame, reducing the visual discontinuities (of dropped –B and –P frames). Moreover P(Marionette) frames are produced between successive I-frames in order to maintain the same frame of normal play and achieve full interactive operations at variable speedups.
The speedups can be derived as follows.

$$SpeedUps = \begin{cases} \dfrac{\mathrm{N}}{\mathrm{N}_{\text{interactive}}}, \quad \mathrm{S_I} = 0 \qquad (FF) \\ S_I x\left(\dfrac{N}{N_{\text{interactive}}} - \dfrac{1}{\Omega}\right) + \dfrac{\mathrm{N}}{\mathrm{N}_{\text{interactive}}}, 1 \le \mathrm{S_I} < \dfrac{\mathrm{TF}}{\mathrm{N}}, (JF) \end{cases}$$

where

- S_I is the number of skipped sequential I-type frames
- $\Omega = \dfrac{RR\_NormalPlay}{N}$

Fig. 7 depicts the number of supported speedups as a function of $N_{\text{int}\,eractive}$ for various numbers of skipped I-type frames.

To bound the size of I -frames of the interactive mode, the encoder uses two predefined (upper-lower) thresholds.

$$Threshold^{upper} = I_{average}(S_{upper}), \quad 0 < S_{upper} \le 1$$

$$Threshold^{lower} = I_{average}(S_{lower}), \quad 0 < \mathrm{S_{lower}} \le 1$$

An I-frame is re-encoded such that its size is between

$$Threshold^{lower} \le I_{bits\_Size} \le Threshold^{Upper}$$

Computer simulation can be used in order to define the values of the two predefined thresholds. Note that the selected values depend on the type of motion of the original video movie (little motion, normal motion, complex motion). In order to achieve minimum additional storage/network bandwidth and acceptable visual quality during the interactive mode the

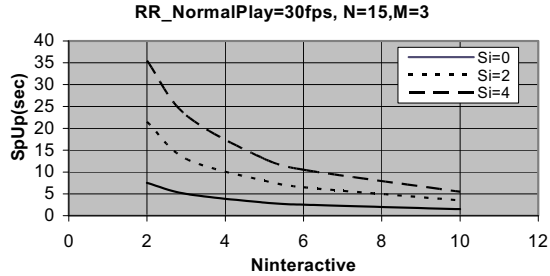


Fig. 7. Relative increase in the speedups as a function of $N_{\text{int}\,eractive}$ for various number of skipped I-frames (S_I).

following values of the predefined thresholds have been selected for the mobile video sequence.

$$Threshold^{upper} = I_{average}$$

$$Threshold^{lower} = (0.9) \times I_{average}$$

After a frame of a scan version has been re-encoded as an I-frame the encoding algorithm checks whether the size of the compressed frame is between the two pre-defined thresholds. If it is not, then the quantization factor for the corresponding frame is increased and the frame is re-encoded [15], [16].

By proper encoding media streams {I-P(Marionette frames)}of the original video sequence, interactive functionality can be supported with considerably reduced additional storage, network bandwidth and decoder complexity and acceptable visual quality at the clients end.

III. Comparative Study

Table 1 summarizes the various aspects of each Fast Forward/Fast Rewind method. Note that none of the previous methods fully address the problem of supporting interactive operations with minimum additional resources. In addition they support limited interactive functions.

We compare our approach to the following approaches (a) Send only I- P- frames [9], (b) Group Of Pictures Skipping [10], (c) Partial Group Of Pictures Skipping[11], (d) Skipping raw of frames [12], Alternative special file [13], I-P(Marionette) frames [14].
The comparison is performed with respect to the following factors

- Additional Resources refers to server load/network bandwidth and decoder complexity
- Functionality refers to the range of the interactivity. Full range of interactive operations support the following functions Fast/Jump Forward, Fast/Jump Rewind, Slow Reverse/|Forward, and Reverse.
- Visual Quality refers to the subject and objective quality of displayed video during the interactive mode

Version	Additional Resources	Functionality	Visual Quality
Send only I - P frames	Significant increase in the bit rate	Limited	Possible problems with Fast Reverse
GOP-Skipping	No increase if 1GoP/disk	Any	Problem with open GoPs
Partial GoP-skipping	Increase in load server/ network bandwidth	Any	Possible problems with Fast Reverse
Skipping raw of frames	Significant extra storage at the server	1 per file	Increase decoder complexity for FR
Alternative special File	Extra Storage at the server	1 per file	Variable consumption rate at the decoder
I-P(Marionette)	No	Any	Acceptable

Table 1: Comparative Study of the proposed encoding media algorithms

The comparison is only meant to convey the tradeoffs provided by different schemes. Table 1 depicts that only the I-P(M) encoding algorithm supports full range of interactive functions with minimum additional resources at the server load, network bandwidth and decoder complexity. In addition all the other approaches support limited interactive functions.

IV. CONCLUSIONS

An MPEG coded media sequence is typically partitioned into small intervals called GoP (Group Of Pictures). This structure allows a simple realization of forward (normal)-pay operation but imposes several additional constraints on the other interactive functions (Fast/Jump Forward, Fast/Jump Rewind, Slow Reverse/|Forward, and Reverse). In this paper, we investigated the constraints of supporting interactive media streaming with respect to Additional Resources, Functionality and Visual Quality of the proposed algorithms. A comparative study has been also carried out in order to determine the tradeoffs between the different methods. Future work involves network simulation of all the algorithms over combined networks from wireline to wireless links

REFERENCES

[1] M. Etoh and T. Yoshimura, “Advances in wireless video delivery,” Proceedings of the IEEE, vol. 93, no. 1, pp. 111–122, 2005.

[2] D. -P. Wu, Y. -T. Hou, W. Zhu, Y.-Q. Zhang, and J. Peha, “Streaming Video over the Internet: Approaches and Directions,” IEEE Trans on Circuits and Systems for Video Technology, Feb 2001, pp 1-20.

[3] G. J. Conklin, G. S. Greenbaum, K. O. Lillevold, A. F. Lippman, and Y. A. Reznik, “Video Coding for streaming media delivery on the Internet, “, IEEE Trans. Circuits Syst. Video Technol., Mar 2001, pp 269-281.

[4] N. Cranley, L. Fiard and L. Murphy, “Quality of Service for Streamed Multimedia over the Internet”, Proc. Irish Signals and Systems Conference 2000, Dublin, Ireland, June 2000

[5] H. Song and C.-C. Jay Kuo, “A region-based H.263+ codec and its rate control for low VBR video,” IEEE Trans. Multimedia, vol. 6, no. 3, pp. 489–500, Jun. 2004.

[6] Coding of Moving Pictures and Associated Audio.MPEG98/W21994, (MPEG-4), Mar.1998.

[7] ITU-T, Recommendation H.263 : Video Video Coding for Low Bit Rate Communications, version 2, March 1993

[8] MPEG video group. Information Technology, Generic coding of moving pictures and associated audio, ISO/IEC 13818-2, International standard, 1995

[9] HJ Chen, A. Krishnamurthy, TDC Little, and D. Venkatesh, “A Scalable Video-on Demand Service for the Provision of VCR-Like Functions,” Proceedings IEEE Multimedia, 1995, p. 65-71.

[10] M-S. Chen, D.D. Kandlur, and P.S. Yin, “Support for fully interactive playout in a disk-array-based video server. In Proc of Second International Conference on Multimedia, pp 391-398, Oct 1994.

[11] Banu Ozden, Alexandros Biliris, Rajeen Rastogi, Avi Silberschatz. “A Low-Cost storage Server for Movie on Demand Databases”. In: Proc. of the 20th VLD Conference, Santiago, 1994.pp 594-605..

[12] Marwan Krunz, George Apostolopoulos. “Efficient Support for interactive scanning operations in MPEG-based video on video on demand”, Multimedia Systems, vol.8, no.1, Jan. 2000, pp.20-36

[13] Michael Vernick, Chitra Venkatramani, Tzi-Cher Chinueh “Adventure in Building the Stony Brook Video Server”. In Proc ACM Multimedia, Boston, Nov 1996, pp 287-295.

[14] Kostas. E. Psannis, M. G. Hadjinicolaou, and A. Krikelis, “MPEG-2 streaming of full interactive content,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 2, pp. 280–285, 2006

[15] Kostas Psannis, Marios Hadjinicolaou and Yutaka Ishibashi, , Efficient Support of Wireless Video Multicast Services in 3G and Beyond, IEEE International Conference on Telecommunication and Networking (http://www.cisse2005.org/), December 2005, pp 262-272

[16] Kostas Psannis, Marios Hadjinicolaou and Yutaka Ishibashi, Efficient Support of Wireless Video Multicast Services in 3G and Beyond, Advances in Computer, Information, and Systems Sciences, and Engineering, (Eds.) Elleithy, K.; Sobh, T.; Mahmood, A.; Iskander, M.; Karim, M. Springer Signals and Communications , Hardcover, ISBN-13: 978-1-4020-5260-6 (October 2006)

The Adaptive Potential of Reconfigurable MEMS in MIMO Antenna Technology

Ligia Chira Cremene, Nicolae Crisan
Communications Department
Technical University of Cluj-Napoca
Romania
Ligia.Cremene@com.utcluj.ro, Nicolae.Crisan@com.utcluj.ro

Abstract— **Part of a detailed study of adaptive radio techniques, the work presented in this paper focuses on the potential of MEMS (Micro-Electro-Mechanical Systems) in reconfigurable antenna array technology. After identifying the main challenges and benefits of multiple antenna systems, we have analyzed the four mechanisms in which a multiple antenna system can improve upon the throughput of a traditional wireless network: beamforming, beam-steering, transmit and receive diversity, and spatial multiplexing. In this paper we propose a unified technology for both the MEMS and the metallic pixel-patches that form the reconfigurable antenna array, thus aiming at a less expensive one, that can be unitarily controlled, as a single entity. This will in turn reduce the computational complexity of the control module. We have designed, simulated and analyzed reconfigurable MEMS antenna arrays at 40 GHz, 2.4 and 5 GHz. Reconfigurability of the operating frequency and of polarization are discussed. Based on our observations we can set the premises for new adaptive strategies.**

Keywords - MEMS (Micro-Electro-Mechanical Systems), reconfigurability, MIMO - multiple antenna systems

I. INTRODUCTION

In order to allow wireless communication transceivers to operate in a multiparametric continuously changing environment, reconfigurable adaptive techniques that adjust the structure and parameters of transceivers and achieve the best possible performance in a variety of scenarios need to be devised [1].

Novel approaches have been proposed recently that achieve reconfigurability by introducing parameterization in the transceiver design with respect to the parameters against which reconfiguration is to be performed, such as antenna correlation and CSI (Channel State Information) reliability [2].

Reference [1] presents a short survey of the latest trends and discusses future directions in the area of smart antennas. An overview of the benefits of and most recent advances in smart antenna transceiver architecture is given first. Then the most important trends in the adoption of smart antennas in future systems are presented, such as reconfigurability to varying channel propagation and network conditions, cross-layer optimization, and multi-user diversity, as well as challenges such as the design of a suitable simulation methodology and the accurate modeling of channel characteristics, interference, and implementation losses.

The adoption of smart antenna techniques in future wireless systems is expected to have a significant impact on the efficient use of the spectrum, the minimization of the cost of establishing new wireless networks, the optimization of service quality, and realization of transparent operation across multi-technology wireless networks. Our goals include: achieving best use of limited and costly bandwidth, and electromagnetic pollution mitigation by enabling the high spectral efficiencies offered by multiple antenna systems.

The success of such technologies relies on two considerations that have been often overlooked when investigating smart antenna technologies [1]: first, the smart antennas features need to be considered early in the design phase of future systems (top-down compatibility); second, a realistic performance evaluation of smart antenna techniques needs to be performed according to the critical parameters associated with future systems requirements (bottom-up feasibility).

In [9] reconfigurability of the operating frequency of the microstrip patch antenna is achieved by loading it with a coplanar waveguide (CPW) stub on which variable MEMS capacitors are placed periodically.

Until now, in the design of second- and third generation wireless systems, smart antenna capability was considered an add-on feature, and optimization of the trade-off between complexity / cost and performance enhancements was not performed during the design phase [1]. Adoption of smart antenna techniques in future-generation wireless systems would require the smart antenna feature to be an inherent part of the system design in order to provide the expected beneficial impact on efficient use of the spectrum, minimization of the cost of establishing new wireless networks, enhancement of the quality of service, and realization of reconfigurable, robust, and transparent operation across multi-technology wireless networks. To this end current research effort in the area is focusing on the following critical issues: 1) The design and development of advanced smart antenna processing algorithms that allow adaptation to varying propagation and network conditions and robustness against network impairments, 2) The design and development of innovative smart antenna strategies for optimization of performance at the system level and transparent operation across different wireless systems and platforms, 3) Realistic

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 421–426.

performance evaluation of the proposed algorithms and strategies, based on the formulation of accurate channel and interference models, and the introduction of suitable performance metrics and simulation methodologies, 4) Analysis of the implementation, complexity, and cost efficiency issues involved in realization of the proposed smart antenna techniques for future-generation wireless systems [1].

II. GOALS AND BENEFITS OF RECONFIGURABLE MULTIPLE ANTENNA SYSTEMS

References like [1], [2], [3], [4] highlight the main challenges and benefits of multiple antenna systems, especially when they are smart or reconfigurable. Maximization of spectral efficiency, of SINR (Signal to Interference and Noise Ratio), of diversity and throughput, of system capacity, lower power requirements and (amplifier) cost reduction are some of them.

There are four mechanisms (table I) in which a multiple antenna system can improve upon the throughput of a traditional wireless network, based on single antennas:

- Beamforming and beam-steering,
- Transmit and/or receive diversity
- Spatial multiplexing

Beamforming involves the weighting and summation of the signals at each antenna to control the angular response of the array. With knowledge of the channel, this technique can be applied to the transmit array as well as the receive array. Beamforming can extend the range of a WLAN (Wireless Local Area Network) by increasing the SNR at the receiver; and higher throughputs may be possible due to the enhanced link quality. Perhaps one of the most important uses of beamforming is in the angular segregation of the wanted signal from any interferers (either different users on the same network or interference from other networks) [3].

Diversity combining is an established technique for improving the mean throughput of a wireless system by mitigating fading. For example, an information bit can be transmitted simultaneously from two antennas (linked by some form of coding), and then the signals can be combined coherently at the receiver. If one of the spatial subchannels experiences a deep fade, it may be possible to recover the information from the signal on the other spatial subchannel.

For each additional diversity branch, a new sample is made of the probability distribution that governs the fading process, and the chance of the combined signals being badly attenuated decreases. In practice, worthwhile improvements diminish above approximately fourth-order diversity. It is therefore best to use only a few maximally decorrelated diversity branches, by judicious construction of replicated signal information in the space, time, and frequency domains. Indirectly, diversity combining can also enable operation in a higher rate mode because of the greater resilience to fading. [3].

Spatial multiplexing is potentially the most powerful MIMO (Multiple Input Multiple Output) strategy for increasing the peak throughput of a system. The basic philosophy is to simultaneously transmit independent data streams from the different antennas and use multiple receive antennas to separate the interfering streams. In theory, this provides linear growth of the capacity with min(M, N). Good performance is possible when using an optimum maximum likelihood (ML) detector, as this provides Nth order diversity by virtue of N independently received samples of the transmitted signals. An optimally designed MIMO WLAN will achieve performance gains from both diversity combining and spatial multiplexing [3].

Alternatively, beamforming can be used to maximally exploit available signal power and mitigate interference. The best solution is to design a system that balances the contributions from these four mechanisms. In practice, this will probably be achieved by defining several different operating modes that trade throughput for robustness, depending on the prevailing channel conditions or requested quality of service (QoS).

TABLE I. MULTIPLE ANTENNA MECHANISMS

Multiple antenna mechanisms	Philosophy	Advantages	Limitations / Disadvantages
Beamforming	-weighting and summation of the signals at each antenna to control the angular response of the array	-extends the range of a WLAN by increasing the SNR at the receiver. -higher throughputs may be possible due to the enhanced link quality -used for angular segregation of the wanted signal from any interferers -can be used to maximally exploit available signal power and mitigate interference	-identical information is transmitted simultaneously from all antennas; hence, the peak throughput of the system can only be improved indirectly by increasing the modulation order -There are practical restrictions on how much the peak throughput can be increased in this fashion [3]
Beam-steering	-transmitting the same signal with different phases, on multiple antennas	-interference cancellation -link quality maximization (minimum BER, maximum data rate)	-limited degree of freedom -complexity -complex channel information is needed
Transmit and/or receive diversity	-simultaneous transmission -coherent Rx combining	-improves the mean throughput of a wireless system by mitigating fading -enables operation in a higher rate mode because of the greater resilience to fading	-in practice, worthwhile improvements diminish above approximately fourth-order diversity -as with beamforming, increasing the modulation order has only limited potential for improving throughput [3]

Spatial multiplexing	-simultaneously transmit independent data streams from the different antennas and use multiple receive antennas to separate the interfering streams	-the most powerful MIMO strategy for increasing the peak throughput of a system -linear growth of the capacity with min(*M*, *N*)	-vulnerable to "self-interference" between the spatial subchannels, especially when the channels are correlated -an ML detector is often prohibitively complex to implement, and reduced complexity alternatives usually exhibit diminished diversity and correspondingly poorer performance[3]

III. MEMS (MICRO-ELECTRO-MECHANICAL SYSTEMS) TECHNOLOGY

The technology was first developed in the 1960s, by selectively removing an amount of silicon from a membrane. This is called bulk MEMS technology [6], being the most mature among silicon micromachining technologies. There are three general approaches to the fabrication of MEMS [7]: surface micromachining, bulk micromachining, and LIGA (lithography, plating, molding). Surface micromachining is a process based on the building up of material layers selectively remaining or removed by continued processing. The bulk of the substrate remains untouched. In bulk micromachining, large portions of the substrate are removed to form the desired structure. Structures with greater heights can be formed because thicker substrates can be used. The bulk micromachining process is a key fabrication method used for MEMS-based photonic switching in the high-growth optical and wireless markets. LIGA processes combine IC lithography and electroplating and molding to obtain depth. Patterns are created in a substrate and then electroplated to create 3D molds. These molds can be used as the final product, or various materials can be injected into them. This process has two advantages. Materials other than silicon can be used (e.g. metal, plastic) and devices with very high aspect ratios can be built.

A wafer-level thin-film encapsulation process has been demonstrated to package radio-frequency (RF) micro-electro-mechanical systems (MEMS) switches in [8]. Electromagnetic simulation and RF test results before and after sealing show minimal RF degradation of switch performance. Before sealing, the insertion loss and isolation at 10 GHz averaged 0.12 and 10.7 dB, respectively. After sealing, the same devices had an average insertion loss and isolation of 0.12 and 10.1 dB, respectively.

In [9] Reconfigurability of the operating frequency of the microstrip patch antenna is achieved by loading it with a coplanar waveguide (CPW) stub on which variable MEMS capacitors are placed periodically. MEMS capacitors are implemented with surface micromachining technology, where a 1-μm thick aluminum structural layer is placed on a glass substrate with a capacitive gap of 1.5 μm. MEMS capacitors are electrostatically actuated with a low tuning voltage in the range of 0–11.9 V. The antenna resonant frequency can continuously be shifted from 16.05 GHz down to 15.75 GHz as the actuation voltage is increased from 0 to 11.9 V.

IV. ANALYSIS OF RECONFIGURABLE MEMS ANTENNA ARRAYS

The proposed reconfigurable pixel-patch antenna architecture consists of many metallic pixel-patches which are interconnected by means of RF MEMS. The antenna is fed along the diagonal axes so that both vertical and horizontal polarizations can be achieved. MEMS work in ON or OFF condition as shown in Fig. 1 (white is OFF). The ON condition links together all colored pixels and MEMS and

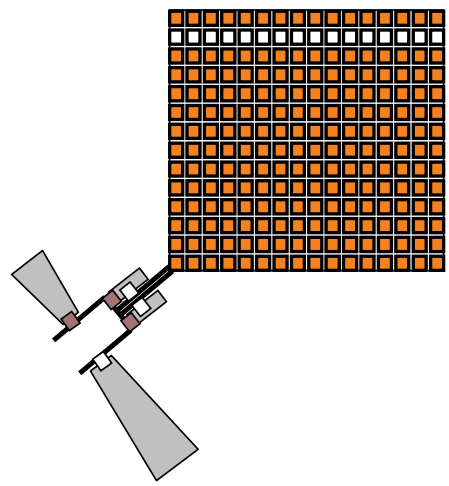


Figure 1. Schematic representation of a reconfigurable MEMS antenna array

leaves the other ones detached (white). The feeding stub adapter is also reconfigurable having two different states suited for reconfigurable MEMS geometry. The wavelengths of the stubs are lengthen up to a quarter-wavelength; with the help of two MEMS (purple in Fig. 1) one stub acts alone, at a certain time, and the another one acts for the other frequency of the reconfigured antenna. The array is designed to operate on two frequencies - 2.4 GHz and 5 GHz - which are selected by switching the proper MEMS actuators.

Fig. 2 illustrates the 2.4 GHz configuration, dual polarization being enabled by means of diagonal feeding (Fig. 3). This is especially useful indoor environment, where wave depolarization is common, due to wall reflections. When the wave arrives with vertical or horizontal polarization (given by the direction of the electric field vector) the antenna can change its polarization in order to adapt its radiation pattern to cope with the incident wave. This ability of the antenna, to

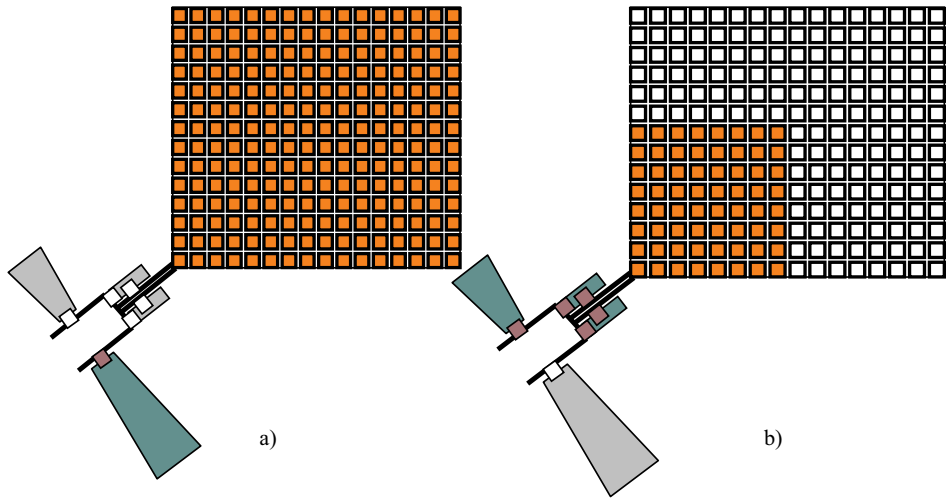


Figure 2. a) The selected configuration for 2.4 GHz (the longer matching stub is active), b) The selected configuration for 5 GHz (the other stub is active and the line is lengthened)

adjust to the right polarization is possible due to a software defined radio architecture acting behind the antenna system. Depending on the operating frequency the matching stub is also reconfigured by adjusting the length of the transmission line.

A series of articles [4], [5], [8], [9] propose some reconfigurable MEMS antennas which are deployed on the monolithic silicon based technology. The classical MEMS technology is different compared to the technology available for the metallic patch manufacturing.

In this paper we propose the use of a unified technology for both the MEMS and the metallic pixel-patches. This idea came from the remark that an antenna system that shares the substrate for both the MEMS and the pixel-patches is less expensive, and can be unitarily controlled as a single entity. This in turn reduces the computational complexity of the control module.

MEMS are deployed on silicon-crystal substrate (which has higher relative permittivity than the silicon) based on a technology which is similar to the microstrip patch technology.

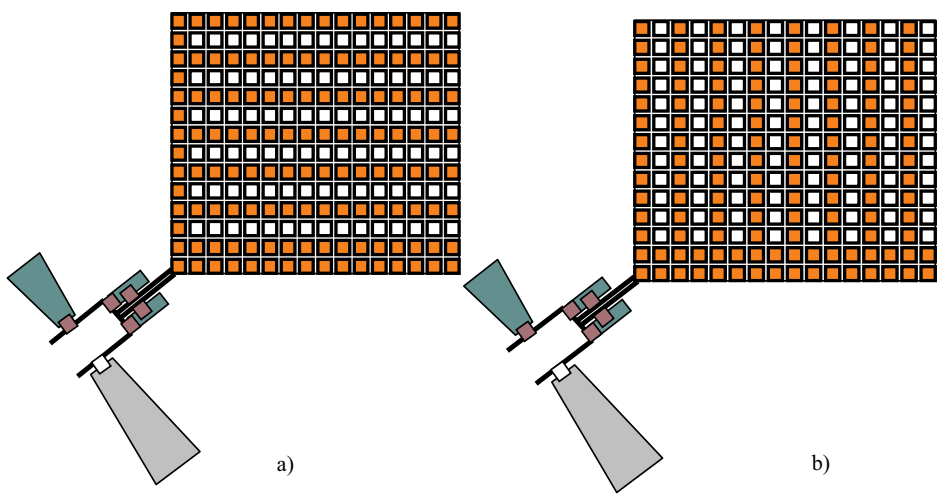


Figure 3. The selected configuration for 2.4 GHz with a) horizontal polarization b) vertical polarization

A problem that we had to deal with was that of matching the MEMS with the pixel-patches, which have different impedance, being designed for 50-ohm RF lines. Changing the MEMS technology is more difficult than changing microstrip dielectric substrate. The larger the pixel-patches the harder is to match the MEMS. So this is the main challenge of the pixelizing process.

When we use the same substrate for both MEMS band patches, if the metallic pixel-patch is much larger than the MEMS, we need a double-end match: pixel-to-MEMS, MEMS-to-pixel. As shown in Fig. 4 the transmission line is placed not in the center of the basic pixel-patch, but closer to a corner, due to matching considerations. Decreasing the dimensions of the pixel-patch causes the transmission line to move towards the center of the patch.

Decreasing the dimensions of the pixel-patch is not possible because the number of MEMS would increase too much, being proportional to the resolution of the patch matrix. We have also proved that matching is still possible even when the size difference between the MEMS and the pixel-patches is significant. In this case we need to adjust the size and position of the transmission line and of the stub (Fig. 4). When different dielectric substrates are used for MEMS and pixel-patches, matching is still possible by adjusting the thickness of the metallic pixels (this is a more expensive technology). In this case the feeding line remains at the center of the pixel-patch, thus preserving the symmetry of the basic unit.

$$W = c / 2f_0 [(\varepsilon_r+1)/2]^{-1/2} \quad (1)$$

$$L = c / 2f_0(\varepsilon_{eff})^{1/2} - 2\Delta l \quad (2)$$

W – patch width
L - patch length
f_0 – operating frequency
h – substrate thickness
ε_r – relative permittivity
ε_{eff} – effective permittivity
Δl – length correction

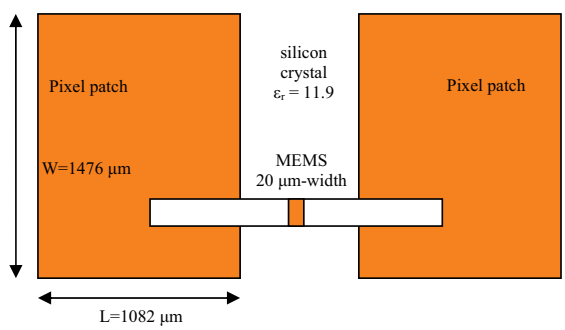


Figure 4. The proposed unified technology for MEMS and pixel-patches: two adjacent, MEMS-interconnected pixel-patches at 40 GHz, all on the same substrate

The antenna array in fig. 5 was designed for the 2.4 GHz band, thus the corresponding array dimensions are L, W = 20mm (formulas (1) and (2)).

During our analysis we observed an interesting phenomenon: when deactivating two border single-patch rows (one horizontal and one vertical), the tuning frequency of the reconfigurable MEMS antenna changes linearly with a small, constant step. In order to decrease the dimensions of the antenna array (Fig. 5) one MEMS line goes in OFF state (on both horizontal and vertical directions), and the corresponding patches are deactivated (white pixels). The frequency step is 22.9 MHz for every consecutive deactivated patch-row. This step was calculated for a VSWR = 1.1 and a corresponding antenna bandwidth of 22.9 MHz. In this experiment the basic dimension of a square pixel-patch is 190 µm. The width of the MEMS antenna array is now 19.81 mm, which includes approximately one thousand pixel-patches (not present in Fig. 5 because of visualizing constraints). Considering the fact that the physical bandwidth for an IEEE 802.11n standard channel is around 22 MHz, the proposed method can be useful for antenna bandwidth reduction with all its related advantages: noise diminishing, better matching and so on.

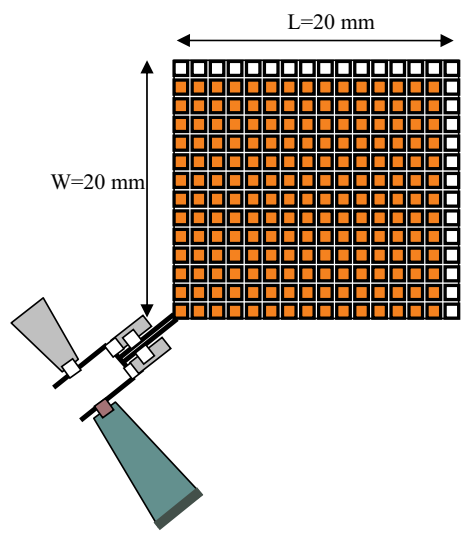


Figure 5. Frequency step tuning for a 2.4GHz MEMS antenna array

V. CONCLUSIONS

We have identified the main challenges and benefits of multiple antenna systems, especially when they are smart or reconfigurable. Maximization of spectral efficiency, of SINR, of diversity and throughput, of system capacity, lower power requirements and (amplifier) cost reduction are some of them.

We have also analyzed the four mechanisms in which a multiple antenna system can improve upon the throughput of a traditional wireless network: beamforming and beam-steering, transmit and/or receive diversity, and spatial multiplexing.

In this paper we propose the use of a unified technology for both MEMS and metallic pixel-patches. This idea came from the remark that an antenna system that shares the substrate for both the MEMS and the pixel-patches is less expensive, and can be unitarily controlled as an entity. This in turn reduces the computational complexity of the control module. A problem that we had to deal with was that of matching the MEMS with the pixel-patches. This was the main challenge of the pixelizing process.

The proposed reconfigurable pixel-patch antenna architecture consists of many metallic pixel-patches which are interconnected by means of RF MEMS. The antenna is fed along the diagonal axes so that both vertical and horizontal polarizations can be achieved. The feeding stub adapter is also reconfigurable having two different states suited for reconfigurable MEMS geometry. The wavelengths of the stubs are lengthened up to a quarter-wavelength with the help of two MEMS. The array is designed to operate on two frequencies - 2.4 GHz and 5 GHz - which are selected by switching the proper MEMS actuators.

During our analysis we have observed an interesting phenomenon: when deactivating two border single-patch rows (one horizontal and one vertical), the tuning frequency of the reconfigurable MEMS antenna changes linearly with a small, constant step of 22.9 MHz. Considering the fact that the physical bandwidth for an IEEE 802.11n standard channel is around 22 MHz, the proposed method can be useful for antenna bandwidth reduction with all its related advantages: noise diminishing, better matching.

We have designed, simulated and analyzed reconfigurable MEMS antenna arrays at 40 GHz, 2.4 and 5 GHz. Reconfigurability of the operating frequency and of polarization were discussed.

REFERENCES

[1] Angeliki Alexion, Martin Haardt, “Smart Antenna Technologies for Future Wireless Systems: Trends and Challenges”, IEEE Communications Magazine, pp. 90-97, 2004

[2] K. Peppas *et al.*, “System Level Performance Evaluation of HSDPA Enhanced through Reconfigurable MIMO Techniques,” IEEE Globecom, Dallas, TX, 29 Nov.–3 Dec. 2004.

[3] Steve Parker, Magnus Sandell, Mong Suan Yee, Yong Sun, Mohamed Ismail, Paul Strauch, and Joe McGeehan, Toshiba Research Europe Ltd., “Space-Time Codes for Future WLANs: Principles, Practice, and Performance”, IEEE Communications Magazine, pp. 96-103, 2004

[4] Bedri Artug Cetiner, Morehead State University, Space Science Center Hamid Jafarkhani, Jiang-Yuan Qian, Hui Jae Yoo, Alfred Grau, and Franco De Flaviis, University of California, Irvine “Multifunctional Reconfigurable MEMS Integrated Antennas for Adaptive MIMO Systems”, IEEE Communications Magazine, pp.62-69, 2004

[5] L. Jofre, Bedri A. Cetiner, F. De Flaviis, “Miniature Multielement Antennas for Diversity Wireless Communications” 2002

[6] Vijak K. Varadan, K.J. Vinoy, K. A. Jose, *RF MEMS and Their Applications*, ISBN 0-470-84308, pp. 51-105, Pennsylvania State University, USA, 2003.

[7] http://www.allaboutmems.com/memstechnology.html

[8] Kevin D. Leedy, Richard E. Strawser, Rebecca Cortez, and John L. (Jack) Ebel, “Thin-Film Encapsulated RF MEMS Switches”, Journal of Microelectromechanical Systems, vol. 16, no. 2, April 2007

[9] Emre Erdil, Kagan Topalli, Mehmet Unlu, Ozlem Aydin Civi, and Tayfun Akin, “Frequency Tunable Microstrip Patch Antenna Using RF MEMS Technology”, IEEE Transactions on Antennas and Propagation, vol. 55, no. 4, April 2007

Voice, Video and Data Transmission over Electrical Power Supply Networks. PLC (Power Line Communications): A Last Mile Alternative for Venezuela.

Luis R. Madera B., Student of Electrical Engineering Mastery, *Central University of Venezuela.*

***Abstract*--In general terms, the public telecommunication networks in Venezuela are conformed by Access, Transport and Core Networks.**

The locations where does not exist telecommunications services but there are users, generally exits electrical services, which allows utilization of existing electrical systems as the access network.

During the development of this research it was consulted documentation related to PLC implementations in Europe, which allowed to establish how is constituted a based PLC network, which are the elements that need to be considered, aspects related with electromagnetic compatibility, and transmission speed researched, also it was consulted the PLC systems description developed by vendors associations and organisms without profit aims, doing more specific in testing PLC systems and results.

This paper presents the research results of information transmission systems over low and medium electrical power supply networks, as a last mile alternative for access to the public telecommunication network in Venezuela using PLC.

I. Introduction

During the last ten years telecommunications systems have grown in an accelerated way and with them the applications available for the end users.

As the TICs evolve, public network must be adapted to offer the requested telecommunications services, which are every time more complex. The emergent services needs bandwidth with more capacity, efficient QoS mechanisms and access technologies every time more flexible, so they can reach more users as possible and fulfill with the telecommunication regulator organism.

This has impelled the need of a constant adjustment of the public network in Venezuela, which in many cases is structured in Access Network, Transport Network and Core Network.

In the Core Network the transference information at great speed and with little manipulation takes place. Generally in this part of the network there are technologies that handle bandwidth up STM1, OC3, Gb, etc.

Transport Network function is the interconnection of different switching devices to exchange information in an intelligent way, between nodes that groups different points of the network, to offer telecommunications services to remotes geographic zones.

The Access Network allows connection to the public network end users, in order to provide the contracted services, through access technologies like: xDSL, ISDN, Frame Relay, ATM, Wireless, etc.

In Venezuela, there are access networks where the external plant conditions do not allows provide telecommunications services with a minimal of quality of service, in others cases the geographic characteristic make the access to network very complicated using a wireless solution, for such situations it is possible use an access network based on PLC systems as a last mile alternative to the access to telecommunications services.

II. Previous Research

The information transmission through the electrical power supply network is not new; previously electrical power supply companies used their network for information transmission at low speed, which allowed implementation of control mechanisms and electrical network management.

The microelectronic evolution, powerful modulation techniques and chip development with more process capacity, it has obtained the creation of devices with great capacity to prepare telecommunications signals that can be transmitted at high speeds through electrical channels.

Many testing have demonstrated that it is possible to send information to rates of transferences that make possible provide traditional and new generation telecommunications services in places where there are only presence of electrical networks, and where the investment required for an infrastructure of external plant makes the access to the information prohibitive.

In countries like Germany, Spain and Malaysia exist experiences in the application of PLC like access technique to

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 427–431.

telecommunications services, also there are indoor type applications where users of old buildings, houses where structured wiring systems do not exist or in new constructions, are using PLC systems as mechanisms for creation of local area networks to share resources, make workgroup and to get access to the Internet from a conventional power outlet.

In Venezuela there are electrical power supply companies making pilot testing to determine the feasibility to use PLC systems to offer telecommunications services, like a value added to the traditional electrical services, this represents a costs reduction to the power supply companies as for end users, because the electrical wiring can be reused for the connection of broadband services in urbanizations, malls, universities, etc.

Although it is certain that PLC systems bring many advantages and benefits, also it is important to consider during the design of an access network based on this system, aspects like: Natural characteristic of transmission channel, electromagnetic compatibility with other existing telecommunications systems, impact of disturbance sources, limitations in the transmission power, performance of MAC layer, topology adjustment of the electrical network and the lack of standards in Latin America that orient the development of the PLC technology according to necessities of the region. These and other factors have diminished the height in the adoption of PLC as solution for the access to the public telecommunications networks in Venezuela.

III. THE PROBLEM

The importance of the access to the information is a subject debated in different world wide scenes; this can be seen in the different reflections made in the UN about the information society, exposition of technological breaches existing between some countries and present consideration of Internet access as an indicator of society development.

In order to satisfy the increasing demand of information, the telecommunications companies are adapting their networks to guarantee the benefit of services with quality in some parts of the national territory, nevertheless exist localities where the access to the telecommunications services is completely null, this is because in many zones of Venezuela an external plant to send and receive information through wired channels does not exist, in others places the geographic situation does not allow using wireless as solution to the access network or the installation cost of a optical access network affect the return of investment in a reasonable time.

The PLC systems allows the establishment of a new access network through the already existing wiring infrastructure, nevertheless it is important to consider during the design phase the creation of mechanisms to obtain an electromagnetic compatibility, because according to previous studies, there are systems affected when others systems are transmitting in the band from 1.6 MHz to 30 MHz, which is used by most of the access systems based on PLC and by PLC Indoor networks.

Another aspect that affects the development of PLC based access networks, is the natural behavior of the electrical wiring, these channels are important sources of disturbance for voice and data since they were designed for the transport of electrical signals and do not for the transport of telecommunications signals, therefore it is necessary access control mechanisms and modulation techniques to guarantee that the signals transmitted can arrive at their destination in reliable form.

The topology of electrical networks can vary in different ways, there are topologies type tree and topologies that connects several low voltage networks to have greater cover, in others zones are observed interconnection of low voltage networks in a structure type ring. These topology changes must be considered during selection of control protocols, access algorithms and control flow, since some of already existing one were developed to operate under a type of network topology, and an incorrect selection can affect the MAC layer performance in PLC networks.

A very important aspect that has contributed with the telecommunications networks and some others technologies is the work made by standardization bodies, in PLC case, there are manufacturers associations that has contributed with the PLC development, also exist many European documents with specifications for the development of PLC systems, nevertheless during this investigation was not possible get initiatives documented in Latin America, oriented to generation of specifications and/or standards about adoption of PLC like a last mile alternative in the region.

IV. METHODS

During the development of this research it was consulted documentation related to PLC implementations in Europe, which allowed to establish how is constituted a based PLC network, which are the elements that need to be considered, aspects related with electromagnetic compatibility, and transmission speed researched, also it was consulted the PLC systems description developed by vendor associations and organisms without profit aims, doing more specific in testing PLC systems and results.

In order to know the evolution of PLC technology it was analyzed documentation developed by organisms in Europe like OPERA. Looking for initiatives oriented to the development of this technology and its implementations. Also it was reviewed vendor associations papers with technical specifications for PLC systems, these institutions have a great amount of information that altogether with other one obtained from some standardization organisms, it was possible know the state-of-the-art of PLC technology in which standards and recommendations it talks about.

In order to verify the aspects related to information transmission capacities, errors detection and correction methods, modulation strategies, access control and services of new generation that supports part of this research; it was analyzed OFDM technical documentation, access systems like CDMA and TDMA, MAC reservation mechanisms, among another one, later it was studied a simulation model for PLC MAC layer.

Once obtained all the information related to the subject of this research, it was classified by interest area, for which the documentation was divided according to the layers of model ISO/OSI. For example it was separated the information that towards reference to physical layer, that treated on the electrical, mechanical and functional characteristic of PLC systems. This was applied up layer three of the ISO/OSI model, considering that in a PLC model of reference could be included the layers of IOS/OSI model where the information flow takes place.

V. Results

During the research it was studied some implementations of information transmission systems over low and medium voltage electrical power supply network.

It is possible send and receive information at speeds over 130 Mbps, within a frequency band between 1,6 MHz and 30 MHz, being therefore in the HF band (High Frequency), also called short wave [1]. Another aspect found was that the mechanisms used to optimize the performance of bit stream in electrical channels are Spread Spectrum and OFDM. The access control technical used in PLC systems was SS-CDMA (Spread Spectrum - Multiple Code Division Access), OFDMA (Orthogonal Frequency Division Multiple Access), OFDMA/TDMA (Orthogonal Frequency Multiple Division Multiple Time Access Division Access), among others. For handling and control of collisions, the CSMA/CA algorithm can be used.

For the establishment of a based PLC access network the use of the following elements is necessary: Base station, which connects a PLC system to its backbone network; PLC modem, whose main function is convert a signal received from PLC network to a standard form that can be processed by conventional communication systems; in some cases the distance between individual subscribers located in the low voltage distribution network and the base station is very long to be connected to a PLC access system, repeaters techniques are required, the repeaters will divided the access network by segments whose distances are supported by the PLC systems. The network segments are separated using different frequency bands or different time slots [2].

There were two ways for subscribers connection to PLC network via conventional connectors, one by direct connection with modems and the other using PLC Gateways, gateways can be placed anywhere in the access network to provide the function of signal regeneration and to divide the PLC access network from the Indoor network.

In order to provide benefit of broadband services, it was observed techniques to send and receive information in full duplex mode, this is obtained using a channel for send information in the user to network way and another for receive information in the network to user way. For example, if the PLC system is using FDMA as media access control and needs information transference in full duplex way, it is used a frequency band for information transmission and another one for information reception, this allows handle greater transmission speeds and therefore the benefit of broadband services.

One ways to evaluate the system answer to different input elements is by systems modeling and simulation, for which inputs models and system model to be evaluated are required, that produces exit signals when it is excited with some of those inputs. In this research was analyzed a generic model of the PLC MAC layer (Fig. 1), because this is the one that performs the main functions in the access to communication channel.

The simulation model analyzed, was developed for the investigation of signaling MAC protocols, and represents an OFDMA/TDMA scheme. There are a number of bidirectional transmission channels that connect network users/subscribers with the base station, which lead to the FDD mode, with symmetric division of data rates between uplink and downlink transmission directions. Duplex modes with asymmetric and dynamic division provide better network performance that fixed mode. However, the investigation considers the MAC protocol for signaling channel, which can be analyzed independently of the applied duplex mode and division strategy [3].

The transmission channel can be accessed by all network stations in the uplink transmission direction while the downlink is controlled by the base station. The subscribers are represented by the network stations that provide multiple telecommunication services (e.g. telephony and Internet). The network stations and base station implement all features of the investigated MAC layer and protocols.

The disturbance for a PLC system can be represented by an ON-OFF model. When it is OFF the channel is disturbed and no transmission is possible, and when it is ON the channel is available. These two states are modeled by two random variables that represent interarrival times and durations of the disturbances. Both random variables are assumed to be negative exponentially distributed.

The duration of a disturbance impulse was set to 100 μs and it was assumed that the noise impulses with duration shorter than 300 μs do not cause transmission errors.

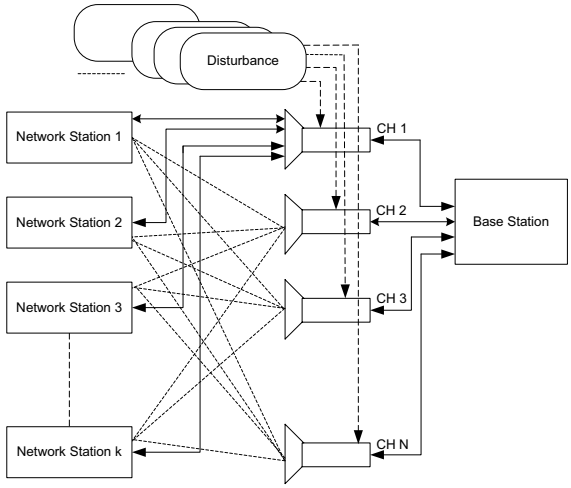


Fig. 1. Generic Simulation Model

The user model can be represented by the network stations providing telecommunications services like data and voice. Both telecommunications services, Internet based data transmission and telephony, representing a packet switched and circuit switched service respectively, was implemented in the simulation model as shown in Fig. 2. The packets (e.g. IP Packet) from the data traffic source are delivered to the packet queue of the network station. Both packet and transmission queue can store exactly one packet. So, a maximum of tow user packets can be stored in the network station. After successful transmission of the packet, the next packet is moved from the packet to the transmission queue. Later the reservation procedure is carried out for the new packet.

The data source generates user packets according applied traffic model. After each packet generation, the data source calculates a time for the generation of a new packet. If the packet queue is occupied, the data source is stopped and it can deliver the new packet after the packet queue is empty again.

The implementation of the telephony service is simpler (Fig. 2). The calls are generated in accordance with a traffic model for the telephony. Generally, for a circuit switched service, the necessary data rate and duration of a connection have to be calculated by the traffic model. The reservation procedure is the same as in the case of the data service. However, the signaling procedure is carried out only once in the case of a circuit switched connection for its setup.

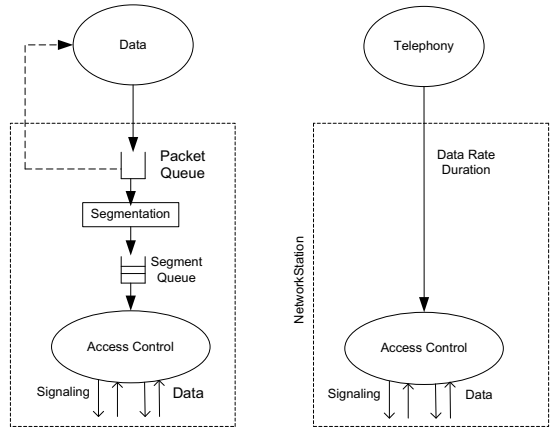


Fig. 2. User Model for Data and Telephony Service

With respect to the quality of service, it was found that the signaling retardation stayed below the 20 ms, which is excellent for data transmission, and can help to guarantee the voice communications, nevertheless is necessary additional use of CAC mechanisms to avoid oversubscription on the network.

The voice transmission can be implemented as a CBR service category, as in a classic telephonic service. It is important to consider the others aspects around the QoS vision when the voice and video are presented on the network.

VI. Discussions

The PLC based access networks are an alternative for the establishment of last mile connections towards the public network, also they can be used for creation of local area networks with the purpose of sharing resources such as Internet connections or access to print servers.

In order to obtain the creation of an access network whose performance guarantees the correct operation for Voice, Video and Data services, it is necessary to consider in first instance the electrical distribution network topology in the low and medium voltage side, it is very important use media access control mechanisms to reduce amount of collisions and the competition by the access to the communication channels. In cases where the distance between the base station and the subscribers is very extensive it is necessary uses PLC Gateways and/or repeaters within network topology with the purpose of improving the network cover.

In order to guarantee a good implementation of PLC based networks it is necessary to know how the different manufacturers solve the subjects related to MAC layer performance, effects of signals disturbance, error detection and correction mechanisms, modulation techniques and the QoS mechanisms at layer 2 level that are used by the devices produced by these vendors and that are placed in the telecommunication market for operators and end users.

With respect to electromagnetic compatibility, it is important that the telecommunications regulation body in Venezuela establishes mechanisms to control the devices power transmission in PLC networks, so these do not generate interference with other already existing devices, without affecting the transmission speed that allows PLC networks offer broadband communication services.

Although this research considered the study of Voice, Video and Data transmission systems over PLC Networks, there was not enough documentation to present the behavior of Video transmission over PLC networks. So, it is necessary to continue the investigation to consolidate the information necessary to know like can the PLC network used for Video transmission, using modeling and simulation systems to support the research in applicable cases to Venezuela involving low and medium voltage networks in the country.

By diverse circumstances in Latin American Countries there are few initiatives for technology generation and standardization, which produces the necessity to acquire for foreign technologies and to adopt standards that are produced outside the region, and in many cases responds to the necessities of the countries working in this developments, which produces vendors dependency, in such sense the mechanisms adoption is important to promote the technology development in the region that allow to acquire not a single technology but a technology package that solves certain necessities of the society while it works by the Latin American technology development.

This investigation is an impulse to continue with the studies and the promotion of the PLC technology, since with the results obtained until now it is possible said that the flexibility and capacities of electrical distribution networks can be taken advantage of no single for electrical transmission and remote protection, but also to obtain the diversification of the access to public telecommunication networks, with the purpose of obtaining information from different connection points.

VII. ACKNOWLEDGMENT

The author gratefully acknowledges the contributions of L. Fernandez, F. Rojo, and R. Ojeda. Thanks so much for their support on this paper.

VIII. REFERENCES

[1] J. R. Gonzales, F. J. Viera, "La tecnologİa PLC en los Programas de Fomento de la Sociedad de la Información de Red.es", Boletin de RedIRIS, nº 68-69, 2004.

[2] H. Hrasnica, A. Haidine, and R. Lehnert, "Broadband Powerline Communication Networks, Network Design". Dresden University of Technology, Germany, 2004.

[3] H. Hrasnica and A Haidine, Modeling MAC layer for Powerline Communications Networks, Internet, Performance and Control of Network Systems, Part of SPIE´s Symposium on Information technologies, Boston, MA, USA, 200.

[4] A. R. Bahai and B. R. Saltzberg, "Multi-Carrier Digital Communications Theory and Applications of OFDM". Algorex, Inc. Iselin, New Jersey, 1999.

[5] J. Reeber and O. Logvinov, "HomePlug & Service Providers: Defining the platform for high-value services in the home". HomePlug Powerline Alliance, 1999.

IX. BIOGRAPHY

Luis Madera was born in Caracas - Venezuela, on June 9, 1975. His employment experience included the Unisys Venezuela, Logimix Sistemas A1 and Damovo Venezuela Companies. His academic experience included the University Institute of Technology "Antonio Jose de Sucre" as Professor of Data Networks, Information Security and Telecommunications Systems and "Alejandro de Humboldt" University as Professor of Teleprocess and Computers Networks.

Luis received degrees as Systems Engineer in December 2000 in the Polytechnic Experimental National University "Luis Caballero Mejias". Post degrees as Specialist in Data Communication Networks in 2004 from UCV University. Current He is student of Electrical Engineering Mastery and Research Management Specialization in UCV University.

Design and Analysis of Optical Interconnection Networks for a Dataflow Parallel Computer

João E. M. Perea Martins Marcos A. Cavenaghi
UNESP – Sao Paulo State University
Department of Computing
17033-360 Bauru, SP – Brazil
E-mail: {perea, marcos}@fc.unesp.br

Abstract- This work shows the design, simulation, and analysis of two optical interconnection networks for a Dataflow parallel computer architecture. To verify the optical interconnection network performance on the Dataflow architecture, we have analyzed the load balancing among the processors during the parallel programs executions. The load balancing is a very important parameter because it is directly associated to the dataflow parallelism degree. This article proves that optical interconnection networks designed with simple optical devices can provide efficiently the dataflow requirements of a high performance communication system.

I. Introduction

This work focuses on the design of optical interconnection networks for the dataflow parallel computer architecture model. The model has an enormous potential for massive data processing, and its model of computation can be associated to a wide variety of computing systems, such as signal processing algorithms, data stream processing, distributing computing workflow, and embedded processing architectures [1, 2].

The construction of dataflow hardware began in the 1970s, and in the 1980s there has been a considerable improvement of this kind of architecture; however, this improvement was not enough to ensure a more realistic dataflow consolidation [3].

Real implementations of dataflow hardware have problems such as the lack of efficiency to hand some specific data structure, and the requirement of a high performance communication system [7]. However, it has advantages as a simple and powerful representation of the parallelism, and data processing directly related to data dependencies.

These advantages ensure a great dataflow potential to solve problems with an enormous parallelism degree, and it justifies new researches about the architecture.

Note that, the dataflow programming execution model has a flexible philosophy and can be used with or without dataflow hardware. This shows that, language improvements aren't necessarily associated to hardware improvements.

In the 1990s the interest in dataflow hardware declined, but due to its potential, the model has never been forgotten. A prove of that is the continuous language evolution that includes visual languages such as the Prograph and the LabView that has a large industrial and commercial focus.

Therefore, the dataflow researches have been more focused on programming languages, and the interest on the dataflow hardware seems stabilized [3, 4]. However, in the last years, due to the development of circuit systems design, the studies and the developments of dataflow hardware turned to embedded systems [2, 5, 6].

These new researches prove the dataflow hardware feasibility, and shows that technological improvement is an important key to make more realistic the dataflow hardware implementation. Therefore, motivated by the technological advancements, this work proposes a dataflow hardware improvement based on optical systems that are also a way to recover the motivation for the development and improvement for parallel dataflow architectures.

The use of optical systems can solve a critical dataflow hardware problem: the requirement of a high performance communication system. Optical high performance interconnection networks have many advantages over electronic systems, such as a theoretical bandwidth of 50 THz [8].

Besides bandwidth, optical networks also implement data format transparency, protocol transparency, passive multiplexing, and electrical noise immunity. Therefore, the optics can be an efficient way to solve critical dataflow problems, and recover the motivation for dataflow hardware research.

Currently, one of the most promising technologies of optical communication is the WDM (*Wavelength Division Multiplexing*) optical coupler. The WDM can multiplex and demultiplex optical signals into one optical fiber; thus, the optical bandwidth is divided into many wavelengths, used as independent communication channels. It is a fundamental key to Gigabits and Terabits networks.

This article introduces the dataflow structure and shows the Wolf dataflow model, in Section 2. Section 3 shows the Wolf simulation. Section 4 describes the optical interconnection simulation parameters used adopted. Section 5 shows the design of two WDM optical interconnection networks based on fixed and tunable optical transmitters. Finally, the work concludes that simple optical systems are efficient to solve the dataflow requirements of high performance communication systems.

II. The Dataflow Architecture And Programming

The dataflow is a kind of parallel computer architecture that involves simultaneous processing of different instructions and data. The operands are assembled in packets, called *tokens*, and programs can be represented by graphs, which structure is

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 432–437.

composed by *nodes* that represent primitive instructions, and by *arcs* that represent the data dependencies between the instructions [3].

The dataflow structure eliminates the complex centralized control and uses the availability of data to fetch instructions rather than the availability of instructions to fetch data, enabling a high parallelism level [2].

Remember that conventional structures based on von Neumann architecture require program counter and global memory access and management, what can be considered a bottleneck [3].

There are different implementations of the dataflow computer architecture. This work uses the *Wolf dataflow architecture*, based on the Manchester Dataflow Machine (MDFM) [9]. Wolf explores variable granularity to achieve a better performance [10, 11], and to overcome data dependency problems observed in the first generation of dataflow machines.

In fact, the Wolf has a simple but efficient structure that can be directly associated to optical network structures. Fig. 1 shows the Wolf dataflow architecture. The *Processor Module* (PM) is responsible for data processing. The PM is configured with N *Processor-Units* (PU) and the *Data Memory Mod*ule (DMM) always sends a token for the first idle PU, using the *Distributing Network* (DN).

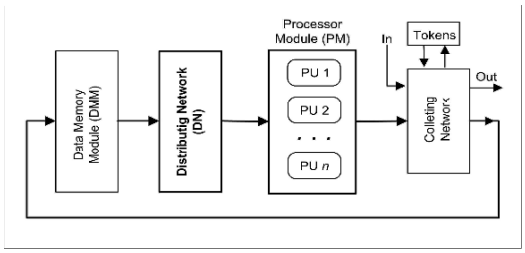


Fig. 1. The Wolf Dataflow architecture.

The Wolf architecture has two different interconnection networks: the Distributing Network and the Colleting Network. The Colleting Network is a more simple system that only controls the flow of information to the Data Memory Module, where it is stored and subsequently distributed for processing.

The Distributing Network controls the flow of data from the Data Memory Module to the Processor Module and, therefore, it supplies information to the Processor Units, and consequently influences directly the processing system operation.

The Distributing Network must be a high performance network to avoid bottlenecks during the program execution, and to ensure a homogenous PU's *load balance*. Therefore, it is an important part of the dataflow hardware, ensuring an efficient execution of the parallel program.

III. Simulation And Analysis Of The Wolf Architecture

This section describes a preliminary simulation and analysis of the Wolf dataflow architecture. The architecture was simulated with a theoretical network using a hypothetical latency of 35 ns. This analysis is based on the utilization of four algorithms:

- Gauss: This algorithm solves an 8x8 linear system of equations using the partial pivoting method,
- Laplace: This algorithm solves a simple Laplace equation using the relaxation method. It uses a 20x20 quadratic matrix,
- Matrix Multiplication: This algorithm generates and multiplies two quadratic matrixes. It is an interactive algorithm and generates a 20x20 matrix as the result,
- FFT: This algorithm implements the Fast Fourier Transform with N = 256 points.

Fig. 2 shows the load balance obtained with the Laplace algorithm simulated in the Wolf simulator using five processing units. The other 3 algorithms (Gauss, Matrix, and FFT) show similar results. The figure reflects an undesirable behavior of the architecture implementation, where the distribution of tokens among the processors is not homogeneous.

This behavior can be caused by the lack of data in the network input, and/or the PUs performance. However, the distributing network performance is a vital key to ensure a satisfactory load balance.

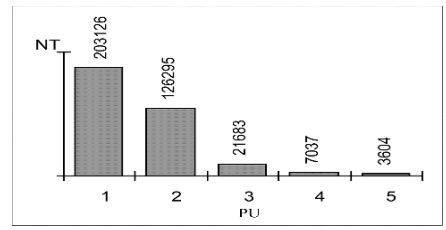


Fig. 2. Load Balancing (Laplace algorithm)

Table 1 shows some parameters achieved from the simulation process, where NTg is the *absolute number of tokens generated* at the Data Memory Module to be sent to the PUs through the Distributing Network, Nsc is the *number of simulation cycles* fixed at 7 ns for simplicity, and Ar is the *Arrival Rate*, representing the amount of data that is generated in the DMM per unit of time.

TABLE 1
Simulation Results

Algorithm	NTg	Nsc	Ar (Gb/s)
Laplace	442,294	4,993,382	3,5
FFT	177,393	1,855,663	3,8
Matrix Multiplication	101,162	1751936	2,3
Gauss	23,690	266220	3,8

Based on the NTg and Nsc, the Ar value is easily computed. We conclude that, for the simulated parameters, DN must have a minimum bandwidth of at least 4,0 Gb/s. However, greater bandwidths are desirable to ensure a satisfactory and efficient Wolf implementation.

This conclusion has motivated the design of optical interconnection networks for this dataflow architecture, whereas the photonic technology can provide efficiently the required bandwidth with the use of simple optical devices.

In the next sections we describe the simulation and analysis of two optical interconnection networks, and prove that the required bandwidth to ensure efficiency in the Wolf dataflow architecture would be at least 10.0 Gb/s.

IV. The Optical Simulation Paramenters

This section explains the simulation parameters that are used in the optical interconnection networks used in the simulation process. This simulation process is focused on the analysis of the optical distributing interconnection network on the load balance.

As stated before, the load balance reflects the token distribution among the several dataflow Processor Units, and determines the amount of processor occupation. Remember that, the number of processors cannot grow infinitely in the parallel computer because the system makes the processing distribution according to parameter such as the language, the architectural model and the interconnection network.

Therefore, the understanding of the load balance is important to ensure real dataflow computer designs with a realistic number of Processor Units. The number of data that are stored in the Data Memory Module is also very important, because its growth can be directly associated to the parallelism level explored in the architecture.

The optical interconnection network simulation process performs a specific *Number of Simulation Cycles* (NSct). Each *simulation cycle* (Sct) lasts a fixed time of 10 ns (Sct=10). The simulation was repeated 10 times, with each simulation process executing 1,000,000 Simulation Cycles (NSct=1,000,000) to ensure good mean values.

The *dataflow token length* (DTL) is fixed at 50 bytes or 400 bits, and the *token processing time* (Tp) required by each PU was computed at 100 ns. When a PU is processing a token, it is treated as a busy device and does not receive another token to process until it finishes processing the current one.

Previously, the Wolf was simulated with 5 PUs, however for optical networks, we used 8 PUs because we suppose that optical networks can improve system performance with the occupation of more processors.

Based on the Ar value, the number of tokens generated per second (NTs) in the DMM is given by (1).

$$NTs = Ar / DTL \quad (1)$$

The time required to send a token (Tk) is given by (2), where *Tr* is the optical transmitter data rate. The *number of tokens* (NTp) that is expected per simulation process can be computed as in (3).

$$Tk = DTL / Tr \quad (2)$$

$$NTp = NTs * NSct * Sct \quad (3)$$

The generation of tokens in the DMM uses a model that is exemplified in Fig 3. During an interval time *dt*, only one packet can be generated in the DMM. This occurs with the probability λdt, independent of other generations outside the time interval.

For a specific interval from t1 to t2, there is a number of arrivals N(t1, t2) that are discretely distributed in the interval, leading to a Poisson distribution model.

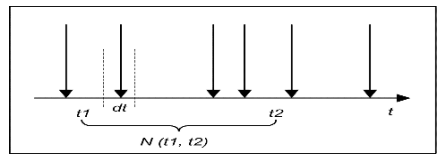


Fig.3. The tokens generation model.

Note that, the Poisson is an important model for packet generation and enables realistic approaches for network simulations. It determines the distribution of the number of data packets arriving for transmission, according to the equation (4) [13].

$$P(k;\lambda) = (e^{-\lambda} * \lambda^k)/ k! \quad (4)$$

Where P(k; λ) is the generation probability of *k* tokens during a period, and λ is the mean number of generated tokens in the period.

V. The Optical Distributing Network Design

The DMM sends the token to the first idle PU. After the transmission, the PU starts automatically the token processing, and simultaneously the DMM starts to send a new token to the next idle PU.

Considering a preliminary proposal of an optical DN model designed with only tunable transmitters, the system operation depends on three basic parameters that are: the *transmitter tuning time* (Tt), the *token transmission time* (Tk), and the *token processing time* (Tp). The Communication Time (Tc) is the sum of Tn and Tk (the total time required to send a token).

This optical DN structure sends tokens only to two PUs, when Tp<Tc. Therefore, the design of a optical DMM depends on the relation between Tp and Tc, because it influences directly the characteristics of the optical transmitters.

This network was simulated for theoretical purposes, with specific parameters. It used only tunable optical transmitters with Tn=1ns and transmitter data rate Tr=2.0Gb/s. From (2), Tk=200ns with Ar values fixed at 2.0 Gb/s.

Simulations showed that PUs number 1 and 2 received 24,834 and 24,779 tokens, respectively. This is an expected result, and it proves that there is a homogeneous token distribution among the processors when Ar and Tc are related.

The simulation process generates 49,613 tokens that is only 0,77% smaller than the theoretical expected value computed from equation 3 (50,000 tokes), what contribute to validate the simulation process. Fig. 4 shows the interconnection network model simulated.

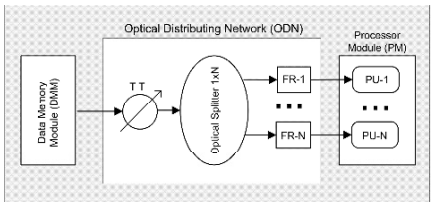


Fig. 4. Preliminary optical interconnection network.

A. The Tunable Optical Distributing Network (T-ODN)

Based on the explanations above, this work proposes the design of an optical interconnection network with tunable transmitters. Fig. 5 shows this network, called *Tunable Optical Distributing Network* (T-ODN). It uses four optical tunable transmitters with Tr=2.0Gb/s.

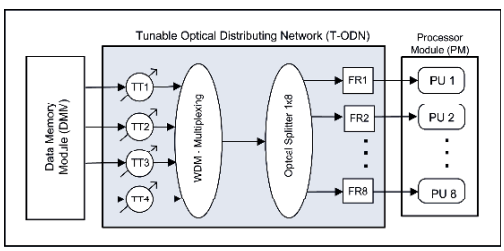


Fig. 5. The Tunable Optical Distributing Network (T-ODN)

Each transmitter can send tokens to two PUs, therefore the T-ODN can send tokens to eight PUs with a bandwidth of 8,0Gb/s, greater than the rate computed in Section 3. However, we must consider the influence of the transmitter tuning time (Tt) on the T-ODN bandwidth (BW). This parameter is computed as in (5), where NTT is the number of tunable transmitters.

$$BW = [Lt / (Tk + Tt)] * NTT \quad (5)$$

Fig. 6 shows the computed bandwidth as a function of the transmitter tuning time (Tt) variation from 0 to 100 ns. For Tt=0ns ensures a bandwidth of 8,0 Gb/s, but for the worse analyzed case, where the Tt=100ns, the T-ODN bandwidth decreases to 5.3 Gb/s.

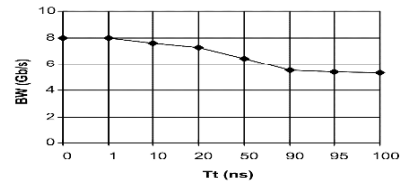


Fig. 6. The bandwidth (BW) as a function of the transmitter tuning time

For a realistic approach, the T-ODN was simulated with a transmitter tuning time equal to or less than 100 ns, and Ar=5.3Gb/s. The simulation also considered Tr=2,0Gb/s, Tp=100ns, and Tk=200ns.

Based on the preliminary considerations at the beginning of this section, it is easy to conclude that for Ar values greater than 5.3 Gb/s, all the PUs will work uniformly, resulting in a homogeneous load balance among all PUs. Despite the desired homogeneous load balance, it was observer that the DMM queue grows.

Anyway, Ar=5,3Gb/s is a significant value and the T-ODN was simulated. Tt was varied, assuming the following results: 1ns, 50ns, and 100ns. Fig. 7 shows the load balance for the values for 100 ns, and Fig. 8 shows the respective DMM queue size variation.

The goal of this analysis was to show the influence of the transmitter tuning time in the optical interconnection network for fixed values of Ar and Tr. We conclude that the optical system can have an efficient operation without the requirement of ultra fast tuning transmitters.

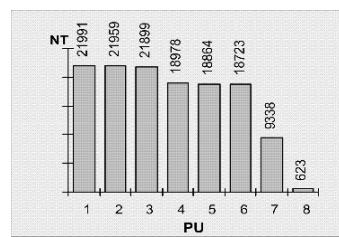


Fig. 7. The number of tokens processed by each processor for Tt=100ns.

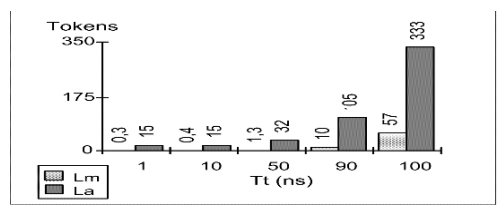


Fig. 8. The number of tokens in the DMM queue, as a function of the tuning time (Tt). *Lm* is the mean value, and *La* is the maximum value measured during the simulation process.

B. The Fixed Optical Distributing Network (F-ODN)

The T-ODN analysis lead to the design of another optical interconnection network for the Wolf dataflow computer, based on the use of fixed optical transmitters, without tuning time.

Fig. 9 shows this optical network. It uses 8 fixed optical transmitters and is named *Fixed Optical Distributing Network* (F-ODN). According to the Mukerjee [14, 15] taxonomy, it is a single-hop FT^8-FR^8 optical network with no transmitter tuning time.

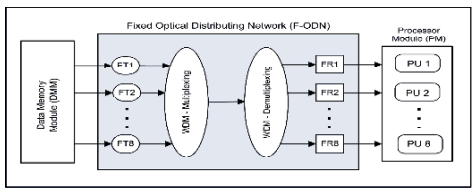


Fig. 9. The Fixed Optical Distributing Network (F-ODN) architecture.

The F-ODN has a bandwidth of 16Gb/s, using eight fixed transmitters with data rate of 2.0Gb/s each. However, for a theoretical study, we consider that a specific processing unit can't receive a new token when it is still processing a previous one.

Therefore, after the transmission of a token, the optical transmitter must wait the ending of the PU processing to start sending a new token. It means that during the PU operation, its optical network channel stay idle, consequently influencing the entire system bandwidth.

Suppose a system with fixed values for Lt, Tk, and Tp. The communication system bandwidth (BW), using *C* optical network channels, is computed as in (6).

$$BW = (Lk / (Tk + Tp) * C \quad (6)$$

Fig. 10 shows that for a theoretical processing time of 0ns, the communication system bandwidth is 16 Gb/s, but it decreases to 10.66Gb/s when the processing time is 100ns, which is a value greater than 4Gb/s computed in Section 3.

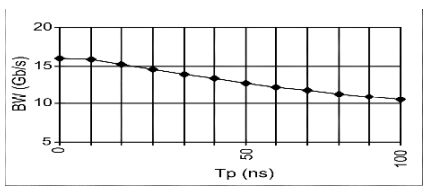


Fig. 10. The communication system bandwidth (BW), considering the influence of the processing time (Tp).

In this work, the F-ODN simulation process was based on transmitter rates of 2.0Gb/s, and processing time of 100ns. The tokens generation rate in the DMM was computed, been in the range from 5.3Gb/s to 10.66Gb/s.

Fig. 11 shows the token queue size at the DMM as a function of Ar. Note that, the queue size growth is proportional to Ar. However, at the worse case, the maximum absolute queue size value is not critical, and it requires a small amount of memory to store the 782 tokens (39.100 bytes).

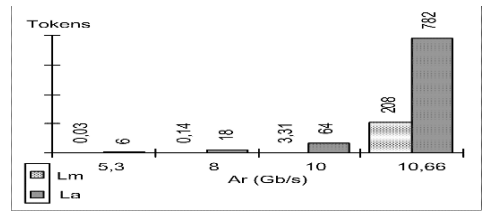


Fig. 11. The F-ODN queue size. *Lm* is the mean value, and *La* is the maximum absolute value measured during the simulation process.

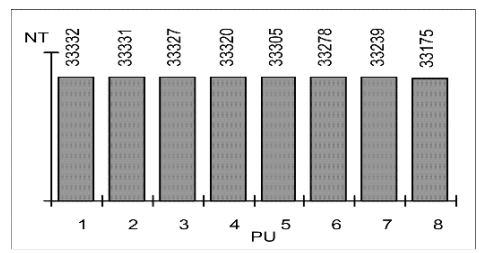


Fig. 12. The Load Balance for Ar=10.66Gb/s.

Fig. 12 shows the load balance for Ar=10.66Gb/s. Another analyzed parameter is the *network throughput* that represents the relationship between the numbers of generated tokens in the DMM and the tokens sent through the F-ODN.

For all values of Ar simulated, the network throughput was greater than 99,9%. This is a very satisfactory value, avoiding big DMM queues, ensuring a high level of PUs occupation, and ensuring a regular load balance.

VI. CONCLUSIONS

The design, analysis, and comparison of two different optical interconnection networks for the dataflow parallel computer architecture have been presented. One can conclude that simple optical systems are very efficient to solve the dataflow requirements of high performance interconnection networks to implement a good load balance.

The first optical network designed with four optical transmitters with data rate of 2,0Gb/s and tuning time of 100ns ensured a bandwidth of 5,3Gb/s, and a satisfactory load balance among six processors of the parallel architecture.

However, the best result was achieved with another optical network design that used eight fixed transmitters with data rate of 2,0Gb/s, ensuring a bandwidth of up to 10.66Gb/s when the data processing time in each processor was 100ns. It ensures both: a satisfactory load balancing among eight processors, and a very small data queue size in DMM.

REFERENCES

[1] S. Neuendorffer, E. Lee: Hierarchical reconfiguration of dataflow models, *Proceedings of the Second ACM and IEEE International Conference on Formal Methods and Models for Co-Design*, June 2004, pp.179–188.

[2] Jackob Carlstrom, T. Bodén: Synchronous Dataflow Architecture for Network Processors, *IEEE Micro*, vol. 24(5), (September/October 2004), pp. 10-18,

[3] Wesley M. Johnston, J. R. P. Hanna, R. J. Millar: Advances in Dataflow Programming languages. *ACM Computing Surveys*, vol. 36(1),: March 2004, pp 1-34.

[4] Dolev Dotan, Ron Y. Pinter: Hyperflow: an Integrated Visual Query and dataflow Language for End-User Information Analysis, *Proceedings of the 2005 IEEE Symposium on Visual Languages and Human-Centric Computing* , 2005.

[5] Jackob Carlstrom et al: A 40Gb/s Network Processor with PISC Dataflow Architecture, *IEEE International Solid-State Circuits Conference*, 2004 .

[6] Yijun Liu, Steve Furber: A Low Power Embedded Dataflow Coprocessor, *Proceedings of the IEEE Computer Society Annual Symposium on VLSI* , 2005.

[7] Toshitsugu Yuba, et al.: Dataflow Computer Development in Japan, ACM SIGARCH Computer Architecture News - *Proceedings of the 4th international conference on Supercomputing*, June 1990, pp. 140-147.

[8] P. P. Mitra; J. B. Stark: Nonlinear Limits to the Information Capacity of Optical Fibre Communications, *Nature*, vol. 411, June 2001, pp.1027-1030.

[9] J. R. Gurd, C. C. Kirkham; I. Watson: The Manchester prototype dataflow computer; *Communications of the ACM*, vol. 28(1), 1985, pp. 34-52.

[10] M. A. Cavenaghi; G. Travieso: A.G. Neto, A simple mechanism to deal with sequential code in dataflow architectures, *IEEE Proceedings - 5th International Conference on High Performance Computing*, 1988, pp. 188-193.

[11] M. A. Cavenaghi, J. A. Martini, A. G. Neto: Proposal to solve workload problems in the simulator of the Wolf architecture. *IEEE Proceedings of the Fourth International Conference/Exhibition on High Performance Computing in the Asia-Pacific Region*, may 2000, pp. 315 – 318.

[12] James Hicks; D . Chiou; Boon Seong Ang; Arving: Performance Studies of Id on the Monsoon Dataflow System. *Journal of Parallel and Distributed Computing*, vol. 18, 1993, pp. 273-300.

[13] E. L. Johnson, K. M. Sivalingam, M. Mishra: Scheduling in Optical WDM Networks Using Hidden Markov Chain Based Trffic Predication, *Photonic Network Communications*, vol. 3(3), 2001, pp. 269-283.

[14] B. Mukherjee: WDM-based local lightwave networks. I. Single-hop systems, *IEEE Network*, vol. 6(3). May 1992, pp.12–27.

[15] B. Mukherjee: WDM-based local lightwave networks. II. Multihop systems, *IEEE Network*, vol. 6(4), July 1992, pp. 20–32.

Tracking of Mobile Nodes in Sensor Networks

Daniel Froß, Jan Langer, Marko Rößler, Ulrich Heinkel

Chair of Circuit and System Design
Chemnitz University of Technology
09126 Chemnitz, Germany
{daniel.fross,jan.langer,marko.roessler,ulrich.heinkel}@etit.tu-chemnitz.de

***Abstract* —We present the development and implementation of an algorithm for tracking the position of a mobile network node. The tracking algorithm uses distance measurements between a mobile node and multiple anchor nodes. The first stage of our algorithm calculates the initial position from distance measurements relative to three (2D) or four (3D) anchors of known position. In a second stage this initial position is improved and tracked over time by feeding a Kalman filter with new distance measurements to neighboring anchor nodes of known position. Both the algorithm and the related communication protocol have been implemented and verified in multiple scenarios by using the event driven OMNET++ network simulator.**

***Index Terms*—localization, tracking, sensor networks, Kalman filter, multilateration**

I. Introduction

NOWADAYS, sensor networks are based more and more on wireless links between single nodes. On one hand, smaller installation effort is pushing this trend. On the other hand, miniaturization in combination with radio based connection allows the placement of nodes in locations which are difficult to access. Sensors on moving parts become possible by a wireless connection. The significance of sensor data usually depends on the position of its acquisition. This becomes even more important when the position needs to be determined dynamically at runtime as for mobile sensors or nodes which are placed to an a priori unknown spot. Using the Global Positioning System (GPS) for this is an insufficient solution due to imprecision, size and price limitations. Positioning based on radio inband distance measurement is a more attractive approach for cheap autonomous nodes requiring low power consumption.

In [1] three algorithms for localizing "position blind" nodes in a wireless multi hop network are discussed and compared. These are *Ad-Hoc positioning, N-hop multilateration* and *Robust positioning*. The proposed approaches are subdivided into a three stage process. In the first stage distances are detected between the blind node and at least three anchor nodes over multiple hops. These values are fused to the unknown node position in a second step. The gained position is optionally elaborated by measurements to single hop neighbors in a third stage. The algorithms were applied to static 2D-networks. Localization and tracking by distance measurements to a moving network node has been considered in [2]. Here a direct intervisibility between anchor and mobile node is presumed.

The system we propose works on the basis of distances between nodes. The distance could either be deduced from a radio signal strength indicator (RSSI) or a signal time of flight. Even methods independent from the radio link could be used to gauge the distance such as radar or ultrasonic systems. From the mobile nodes point of view the communication architectures are classified as being *passive* or *active*. In the first case the mobile node is barely a reflector for the distance measurement. A system of anchor nodes is actively observing the distances to the mobile node and transmitting the values to a central higher level instance that performs the position calculation. In the second case the anchor nodes are used as fixed points signaling their own position to mobile nodes. The mobile nodes conclude their own position from the distance to the anchors.

The system described in this paper supports passive and active architectures. We assume a single hop network for simplification, where only direct connected nodes communicate. The computation tasks are optionally performed on any node in the network by using the local available information. This means that there is no multi hop information passing through the network. Our algorithm is composed of two interfering stages. In the first stage an initial position of the mobile node is computed. This is used as a starting point to estimate the motion of the node by the Kalman-Filter in the second stage. The system returns to the first stage if the estimation process fails due to a lack of satisfactory distance measurements to a minimum of anchor nodes.

II. Initial Position of the Mobile Node

In the two-dimensional case, at least three distinct distances between a mobile node and different anchor nodes are necessary to determine the position of the mobile node unambiguously. Anchor nodes are nodes, which have a fixed

This work was supported in part by the German ministry of Education and Science (BmBF) within the "Innoprofile" initiative.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 438–443.

position that is known to the sensor. To determine the position

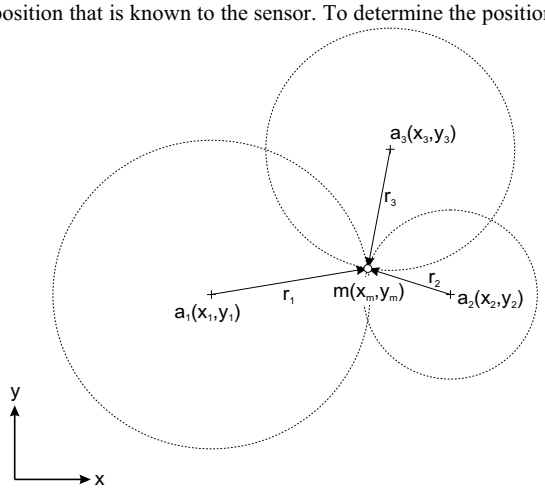


Fig. 1. Position of a mobile node based on three distance measurements to anchor nodes

of the mobile node in the 3-dimensional case, a fourth anchor must be added. However, in this work we consider only the 2-dimensional case.

Fig. 1 shows a scenario of three anchor nodes a_1, a_2 and a_3 together with a mobile node *m* that needs to be located.

In the general case of n anchor nodes, for every anchor $a_i, i \in \{1..n\}$ and distance measurement r_i exists a circle with center coordinates (x_i, y_i) and radius r_i. In absence of measurement errors, the mobile node's position is located exactly on the perimeter of the circle. By considering all measurements, we obtain three circles, which intersect in exactly one point (x_m, y_m) — the position of the mobile node.

Each distance circle can be described as an equation $(x-x_i)^2+(y-y_i)^2=r_i^2$.

By using the equations of all circles, we obtain the following non-linear equation system. Its solution is the point of intersection between the circles and the position of the mobile node.

$$\begin{aligned}&(x-x_1)^2+(y-y_1)^2=r_1^2\\&(x-x_2)^2+(y-y_2)^2=r_2^2\\&\qquad\cdots\\&(x-x_n)^2+(y-y_n)^2=r_n^2\end{aligned} \tag{1}$$

In practical applications the distance measurements and the anchor position contain measurement errors. Therefore, the distance circles do not intersect in exactly one point and the equation system becomes a non-linear optimization problem for x and y. Those systems are usually solved using numerical methods. However, it is often difficult to guarantee numerical stability for those algorithms. Furthermore, the result of numerical algorithms depends on a good starting position.

In [1] and [3] another computation method is proposed. First, the problem is linearized by subtracting one equation from all others. Hence, the non-linear effect of x and y is eliminated. In order to linearize equation system (1), we subtract equation n from all other equations 1…n-1. As a result we obtain the following linear equation system.

$$\begin{aligned}r_i^2-r_n^2=&x_i^2-x_n^2-2(x_i-x_n)x\\&+y_i^2-y_n^2-2(y_i-y_n)y\end{aligned} \tag{2}$$

The equations of (2) can be rearranged and written in matrix notation:

$$A\vec{x}=\vec{b} \tag{3}$$

Matrices A and $\vec{b}$ contain only the measured distance values and the positions of the anchor nodes.

$$A=\begin{bmatrix}2(x_1-x_n) & 2(y_1-y_n)\\ \vdots & \vdots\\ 2(x_{n-1}-x_n) & 2(y_{n-1}-y_n)\end{bmatrix}$$

$$\vec{b}=\begin{bmatrix}x_1^2-x_n^2+y_1^2-y_n^2-r_1^2+r_n^2\\ \vdots\\ x_{n-1}^2-x_n^2+y_{n-1}^2-y_n^2-r_{n-1}^2+r_n^2\end{bmatrix}$$

The result of the linearization step is a system of equations

$$n_{xi}x+n_{yi}y=\left|\vec{n_i}\right|d_{i\perp} \tag{4}$$

with

$$n_{xi}=A_{i1}\quad n_{yi}=A_{i2}\quad \left|\vec{n_i}\right|d_{i\perp}=b_i\,.$$

As shown in Fig. 2, every equation represents a straight line g_{in}. It is orthogonal to the line connecting the positions of anchors a_i and a_n and it runs through both intersection points of the distance circles around both anchors. In equation (4), n_{xi} and n_{yi} are the *x*- and *y*- component of a vector orthogonal to g_{in}, whereas $d_{i\perp}$ is the distance between g_{in} and the origin of the coordinate system.

Without errors, all lines will intersect in one point, which can be easily computed. However, in real systems the equations form a linear optimization problem, which can be solved using the least squares algorithm proposed in [1] and [3].

$$\left\|A\vec{x}-\vec{b}\right\|^2 = \sum_{i=1}^{n-1}(A_{i1}x + A_{i2}y - b_i)^2 \rightarrow \min \qquad (5)$$

An optimal position $\vec{x}_m$ is obtained by solving the corresponding system of normal equations

$\vec{x}_m = (A^T A)^{-1} A^T \vec{b}$.

Please note, that the solution of this linear optimization problem (5) is not the solution of the original non-linear problem. This problem is given by the following formula and represents a least squares error optimization of the actual distances.

$$\sum_{i=1}^{n-1}\left(\sqrt{(x-x_i)^2+(y-y_i)^2} - r_i\right)^2 \rightarrow \min \qquad (6)$$

By rewriting equation (5) as

$$\sum_{i=1}^{n-1}\left(n_{xi}x + n_{yi}y - \left|\vec{n_i}\right| d_{i\perp}\right)^2 = \sum_{i=1}^{n-1}\left(\left|\vec{n_i}\right| \Delta d_{i\perp}\right)^2 \rightarrow \min \qquad (7)$$

we can see that (5) minimizes the scaled squared distances $\Delta d_{i\perp}$ between $\vec{x}$ and each line g_{in}. The corresponding scaling factor is $|\vec{n}_i| = 2\sqrt{(xi - x_n) + (y_i - y_n)}$, i.e. twice the distance between anchor nodes a_i and a_n. Fig. 2 shows the solution of the previous example using the non-linear and the linear optimization problems. In order to see different solutions for both methods, erroneous measurements are used. However, as we use this solution only as the initial position of our tracking algorithm, it is sufficiently correct.

III. Tracking of Mobile Nodes

The tracking algorithm uses a Kalman filter to determine the position of the mobile node. In order to quickly find and correctly track the node a good initial position is necessary. The algorithm provided in section II is used to calculate it.

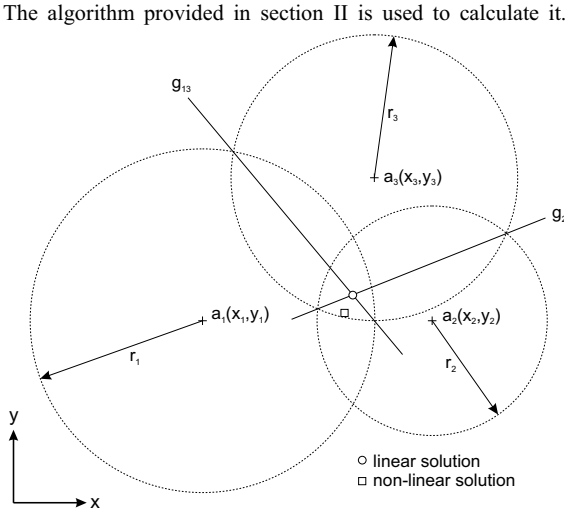


Fig. 2. Solutions of the linear and nonlinear optimization problem

The non-linear dependency between state variables and measurement variables requires the application of an extended Kalman filter. The filter estimates both position and velocity of a mobile node based on a linear motion model with constant speed. In the two-dimensional case, state vector $\vec{x}$ is defined as

$$\vec{x} = \begin{pmatrix} x & v_x & y & v_y \end{pmatrix}^T \qquad (8)$$

where x, y and v_x, v_y represent the x and y component of the node's position and velocity respectively. The filter employs the prediction/correction principle, i.e. each filter update consumes exactly one distance measurement. This filter type is known as SCAAT (Single Constraint At A Time) Kalman filter [4]. Based on an initial state estimate $\hat{x}(0)$ and the estimate covariance matrix $P(0)$, the filter executes the following calculation steps for each single incoming measurement at time t.

1. *Determination of the time difference Δt between current time t and the time of the previous estimation step.*
2. *Prediction of state estimate $\hat{x}^-$ and estimate covariance matrix P^-*

$$\hat{x}^- = A(\Delta t)\hat{x}(t - \Delta t)$$

$$P^- = A(\Delta t)P(t - \Delta t)A^T(\Delta t) + Q(\Delta t)$$

The state transition model $A(\Delta t)$ of the system contains the chosen constant speed motion model.

$$A(\Delta t) = \begin{pmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad (9)$$

Covariance matrix $Q(\Delta t)$ represents the uncertainty of the state prediction. According to [5] it is defined as follows

$$Q(\Delta t) = \begin{pmatrix} Q^*(\Delta t) & 0 \\ 0 & Q^*(\Delta t) \end{pmatrix}$$

$$Q^*(\Delta t) = \begin{pmatrix} \Delta t^2 & \Delta t \\ \Delta t & 1 \end{pmatrix}\sigma_v^2 .$$

The variable σ_v^2 corresponds to the velocity variance that the system assumes.

3. *Prediction of distance estimate $\hat{r}_i$ and calculation of the corresponding Jacobian matrix H.*

$$\hat{r}_i = \sqrt{(\hat{x}^- - x_i)^2 + (\hat{y}^- - y_i)^2} \quad (10)$$

In this equation x_i and y_i represent the anchor node a_i according to distance estimate $\hat{r}_i$. The partial derivatives with respect to the velocity evaluate to 0 since we assume a constant speed model.

$$H = \left(\frac{\partial \hat{r}_i}{\partial \hat{x}^-} \quad \frac{\partial \hat{r}_i}{\partial \hat{v}_x^-} \quad \frac{\partial \hat{r}_i}{\partial \hat{y}^-} \quad \frac{\partial \hat{r}_i}{\partial \hat{v}_y^-} \right) = \left(\frac{\partial \hat{r}_i}{\partial \hat{x}^-} \quad 0 \quad \frac{\partial \hat{r}_i}{\partial \hat{y}^-} \quad 0 \right).$$

The derivatives with respect to the node's position can be calculated as

$$\frac{\partial \hat{r}_i}{\partial \hat{x}^-} = \frac{\hat{x}^- - x_i}{\sqrt{(\hat{x}^- - x_i)^2 + (\hat{y}^- - y_i)^2}} = \frac{\hat{x}^- - x_i}{\hat{r}_i}$$

$$\frac{\partial \hat{r}_i}{\partial \hat{y}^-} = \frac{\hat{y}^- - y_i}{\sqrt{(\hat{x}^- - x_i)^2 + (\hat{y}^- - y_i)^2}} = \frac{\hat{y}^- - y_i}{\hat{r}_i}.$$

4. *Calculation of Kalman gain K.*

$$K = \frac{P^- H^T}{H P^- H^T + R} \quad (11)$$

The Kalman gain is a weighting factor indicating to what degree the predicted system state is adjusted by the actual measurement. Matrix R allows the system to deal with measurement noise. In case of our SCAAT algorithm, which updates the filter with individual measurements it simplifies to a scalar variable. The value of this variable corresponds to the variance of the measurement error, which is assumed to be normally distributed.

$$R = \sigma_{r_i}^2 \quad (12)$$

5. *Calculation of the measurement residual Δr_i indicating the difference between the actual measurement and its orediction.*

$$\Delta r_i = r_i - \hat{r}_i \quad (13)$$

6. *Update of state estimate and estimate covariance.*

$$\hat{x} = \hat{x}^- + K \Delta r_i$$
$$P = (I - KH) P^-.$$

Matrix I is the two-dimensional identity matrix.

In the two-dimensional case, a mobile node must measure the distance to at least three distinct anchor nodes, in order to obtain a determined position. As the filter works on single measurements (SCAAT) it will not detect violations of this condition by itself. The mobile node implements a corresponding strategy, e.g. by switching off the estimation, when too few nodes are available over a certain time window.

IV. Network simulation

Localization and tracking of mobile nodes by using the methods described in section II and III have been implemented in *C* language. The implementation is independent from the communication architecture (active or passive). A passive architecture has been used as a simulation scenario to verify our methods. The simulation was executed in the open source network simulator *OMNeT++* [6].

OMNeT++ is based on a discrete event driven simulation kernel, which is implemented in *C++*. An object oriented framework for modeling networks is delivered with *OMNeT++*. It provides basic components, messages and a blackboard. Special components and messages can be inherited from the basic versions to implement customized functionality. The NED language is then used to compose the network to a widely parametrizable model. The simulator comes with a graphical user interface that allows not only visualization and control of the simulation but also extensive debugging. Modeling wireless networks is not supported natively. The *Mobility Framework* [7] extension developed at Berlin University of Technology can be used for this purpose. It utilizes the strict modular concept of *OMNeT++* to allow simulation of mobile wireless connected network nodes.

The *Mobility Framework* provides a three-tier layer for the mapping of network stacks. Each layer comes with operational example implementations of common implementations. The lowest *Network Interface Controller* layer is used to model the behavior of the physical medium including error correction and medium access control mechanisms. In our model the 802.11 WLAN implementation with *Carrier Sense Multiple Access* (CSMA) mechanism was utilized. Routing algorithms are covered by the next higher *Network* layer. In our single hop network this layer has no function.

The highest *Application* layer is used for the localization functionality. Messages generated on this layer deliver position and distance information between nodes. The *Mobility Framework* comes with various standard motion models, to recreate a nodes movement. They where adapted to closely recreate the natural indoor movement of a human.

The sensor network simulation model consists of an arbitrary number of anchor and mobile nodes as shown in Fig. 3. The entire functionality for the self contained localization and tracking has been encapsulated in a *C* implementation called *node* covered in a shared library *libnode*. At the beginning of a simulation all nodes can be placed randomly or dedicated on the simulation playground. After a randomly chosen period the anchors begin to

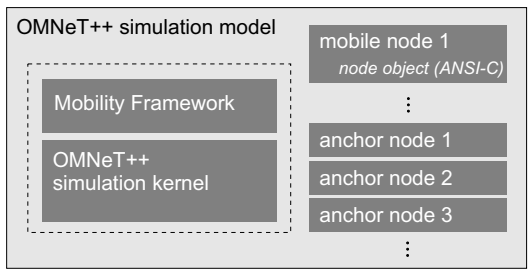


Fig. 3. Simulation model in OMNeT++

cyclically generate broadcast messages containing their real position on the playground. Mobile nodes in range receive these messages and get to know about the anchors existence. A list of neighbors is gathered using this information. A list entry contains the position, the distance information to that anchor and the timestamp of the measurement. The timestamp is local as there is no time synchronization between network nodes. In the simulation model the distance measurement is taken at the time of a broadcast message receipt. A randomly chosen delay is put on the moment of sending the broadcast. This is done to model the fact that in reality the distance measurement takes an indefinite amount of time.

The multilateration stage is executed once the neighborhood list contains more than three entries. A contemporaneously distance measurement is assumed at this point. This would never be the case in a real scenario where the measurement is occupying a shared resource. For any degree of accuracy the difference of the measurement timestamps must not exceed a certain interval of time. This interval depends on the speed of the mobile node.

After a successful multilateration we iterate cyclically over the neighborhood list to feed the estimation stage with current distance information. When new anchor nodes appear in the range of receipt they are automatically inserted in the list. Entries in the list are deleted after a certain period of broadcast inactivity. A further check on the list ensures that a minimal number of neighbors provide distance information in order not to feed the Kalman estimator under constrained. The tracking is canceled if the list contains measurements to less then N anchors (the system is under constrained for $N<3$) within a selectable time window. It is therefore possible to bridge under constrained zones o and therefore track a mobile node in less anchor dense areas. After cancellation, the algorithm returns to stage one when $N \geq 3$ anchors are in range again.

V. Results

Figures 4 to 6 show the simulation results for a position estimation of a mobile node that moves with a constant speed of 3 m/s in the network. The straight line corresponds to the real position of the node. In the best case the mobile node receives distance information every 500 ms from all anchors in communication range. White Gaussian noise of 10m standard deviation has been superimposed on the distance information in order to model measurement errors. The Symbols represent the estimated position of the mobile node. The type of symbol encodes to number of anchors in range at the moment of estimation. The lines connect the symbols as estimated position to the real position of the mobile node in that particular point in time. The length unit in all figures is meter.

Fig. 4 shows the result for our final version of the Kalman filter. The initial position is determined basing on three measurements right at the start of the simulation. From that point the Kalman filter estimates the position as long as at least three anchors are in range. The filter has a sluggish setup in order to optimally cope with measurement errors. On the other hand it is still dynamical enough to follow the course of motion.

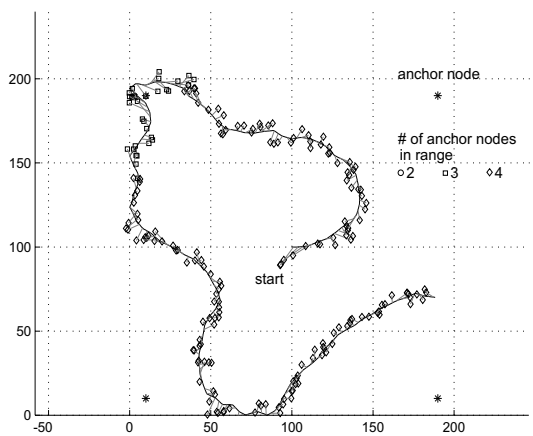


Fig. 4. Kalman filter with motion estimation

This can be compared to Fig. 5 where the same scenario is shown without considering the speed in the position forecast. The estimated position stays permanently behind the real

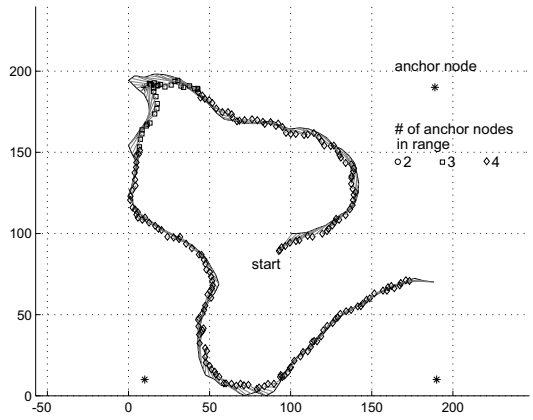


Fig. 5. Kalman filter without motion estimation

course of movement. This behavior is even more distinctive with an increasing speed of the mobile node. The lack of incorrect estimated positions is the major advantage of a filter without motion estimation. Therefore the curve appears to be more balanced compared to Fig. 4.

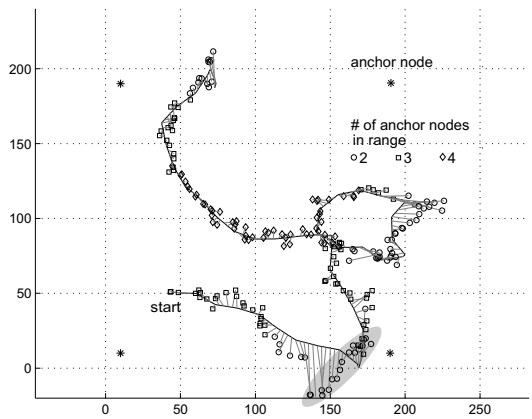


Fig. 6. Kalman filter operates under constrained

Fig. 6 shows the example of an under constrained operation of the estimation filter. Under constraining the Kalman leads to the tracking of an *alias* motion at the marked spot. Therefore the tracking has to be canceled as soon as a permanent lack of diversity in distance information occurs.

VI. Summary

The work presented in this paper covers an algorithm for localization and tracking of mobile network nodes based on distance measurements to multiple anchor nodes. The algorithm and a simulation scenario have been implemented in *C*. The event driven network simulator *OMNeT++* and the *Mobility Framework* extension have been used to verify our approach. Modeling motion and communication of the nodes allowed us to qualify appropriate parameter settings of the algorithm in different scenarios. In the next step we plan to integrate our approach in real microcontroller driven sensor nodes that support inband distance measurement. Further work concentrates on extending our methods to multi hop networks in order to push the precision. Integrating position and distance information of nodes in two or tree hops distance leads to more accurate positioning. Mobile nodes with contact to less then three anchors will than are able to get located.

References

[1] K. Langendoen and N. Reijers, “Distributed localization in wireless sensor networks: a quantitative comparison,” *Computer Networks*, vol. 43, no. 4, pp. 499–518, 2003.

[2] A. Smith, H. Balakrishnan, M. Goraczko, and N. B. Priyantha, “Tracking Moving Devices with the Cricket Location System,” in *2nd International Conference on Mobile Systems, Applications and Services (Mobisys 2004)*, Boston, MA, June 2004.

[3] A. K. L. Miu, “Design and Implementation of an Indoor Mobile Navigation System,” Master’s thesis, Massachusetts Institute of Technology, Cambridge, MA., Januar 2002.

[4] G. Welch and G. Bishop, “Scaat: incremental tracking with incomplete information,” in *SIGGRAPH ’97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*.1em plus 0.5em minus 0.4emNew York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1997, pp. 333–344.

[5] D. C. Moore, “Robust Distributed Sensor Network Localization with Noisy Range Measurements,” Master’s thesis, Massachusetts Institute of Technology, Cambridge, MA., Februar 2005.

[6] A. Varga, “The omnet++ discrete event simulation system,” in *European Simulation Multiconference (ESM’2001)*, June 2001. [Online]. Available: http://www.omnetpp.org/download/docs/papers/esm2001-meth48.pdf

[7] W. Drytkiewicz, S. Sroka, V. Handziski, A. Koepke, and H. Karl, “A mobility framework for omnet++,” 3rd International OMNeT++ Workshop, at Budapest University of Technology and Economics, Department of Telecommunications Budapest, Hungary, Jan. 2003.

IP Based Mobility Management for Next Generation Wireless Networks

Md. Akbar Hossain*, Khan Md. Rezaul Hoque†
*Department of Telecommunication Engineering, Faculty of Engineering
† Department of Information and Communication Technology, Faculty of Science
University of Trento, Italy
Email: mdakbar.hossain@studenti.unitn.it, khan@dit.unitn.it

***Abstract*- During the last decade wireless systems has experienced tremendous demands from social market. Subscribers now want a wider choice of services, better quality, faster response times, and greater coverage, which enhance the development of wireless technology than ever before. Among these, one of the greatest concerning problems for seamless global connectivity of mobile user is mobility management. The aim of this paper is to analyze the IP based Mobility management for next generation wireless networks. By the way of mobility management it is possible to locate roaming terminals for call delivery and maintain connections with mobile terminal that changes their points of attachment. In this paper it is also discussed how several type of mobility can be solved by Mobile IP and cellular IP.**

I. INTRODUCTION

Now a days wireless communication systems are the fastest growing area of the communication sector. The protocol of IP plays an important role in the field of mobility management in various types of wireless networks. One of the key features of 4G is that it will be based on an all-IP infrastructure for both fixed and mobile networks. Moreover, positioning on network layer, IP acts as a masking isolator that prevents the protocols, services, and applications of upper layers from the awareness of network interconnecting architecture and possible changes caused. So, as a suitable layer to solve the problem of mobility and provide transparent mobility to applications and higher level protocols like TCP, IP becomes one of the most important research issues in mobility and location management, which results in various techniques and standards based on the extension of fundamental IP protocol proposed. The main goal of the mobile IP working group is to develop routing support to permit IP nodes (hosts and routers) using either IPv4 or IPv6 to seamlessly "roam" among IP sub networks and media types.

This article is organized as a general description and operation of mobile IP and IP based mobility management in section 2. In section 3 mobile IP optimization and advantages of MIPV6 are also discussed. Intra-Domain Mobility management is described in section 4. Cellular IP as a solution of mobility management for next generation wireless communication is also describe in section 5.

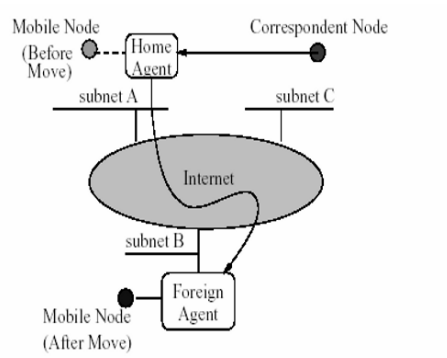


Fig. 1 Architecture of Mobile IP

II. IP BASED MOBILITY MANAGEMENT

A simple Mobile IP architecture is illustrated in Fig. 1. In this example, the CN sends packets to the MN via the MN's HA and the FA.IP-based networking is designed such that each host is identified by a unique IP address. Standard IP routing assumes that IP addresses are distributed hierarchically. For example, a host with a certain subnet prefix is assumed to be located at the subnet referenced by that prefix, the home network. This dual use of IP addresses is fine when hosts are not mobile, as each host can be assigned its unique IP address according to the hierarchical structure needed for IP routing. However, it creates a problem when hosts need to be mobile. If a host moves to a foreign network, packets for it will still be routed to its home network. Furthermore, a host may obtain a temporary address in the foreign network for routing purposes, but there is no

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 444–447.

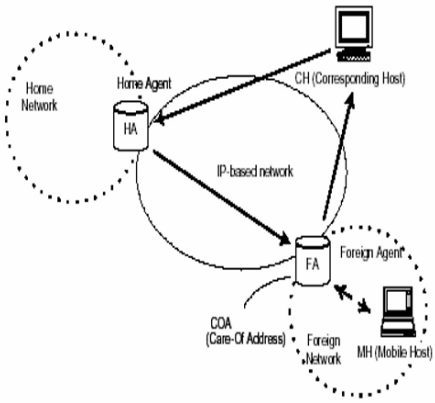


Fig. 2. Mobile IP with foreign agent

association between its temporary and permanent addresses. In MIP, each Mobile Host (MH) is still identified by its permanent IP address. However, for routing purposes, when an MH is roaming it obtains a temporary care-of-address (COA), which is a foreign network address that identifies the location of the MH. The MH registers this COA with a mobility agent in its home network known as its Home Agent (HA). The HA then stores the COA of the MH in a binding cache. Nodes communicating with the MH send packets addressed to its permanent address. These packets are routed to the MH's home network, where it's HA intercepts them and tunnels them (encapsulated) to its COA. The MH registers its latest COA with it's HA whenever its COA is changes, which occurs when the MH moves to another foreign network. It should also refreshes the registration with it's HA periodically. MIP can operate in two modes, namely with foreign agents or with co-located COAs, illustrated respectively in Fig. 2 and Fig. 3.

In the mode with foreign agents, the visited network has a Foreign Agent (FA). The FA broadcasts its IP addresses that can be used as COAs. The MH picks a valid IP address of the FA as its COA and registers this with it's HA (in this mode, the registration goes through the FA rather than directly to and from the HA). When packets arrive for the MH at the FA tunneled from the HA, they are un-encapsulated and forwarded to the MH through its layer 2 address previously registered with the FA. On the other hand, in the mode with co-located COAs, the MH would obtain a temporary IP address at the foreign network using a protocol such as DHCP (Dynamic Host Configuration Protocol). The MH would use this temporary IP address as its COA and registers this with HA.

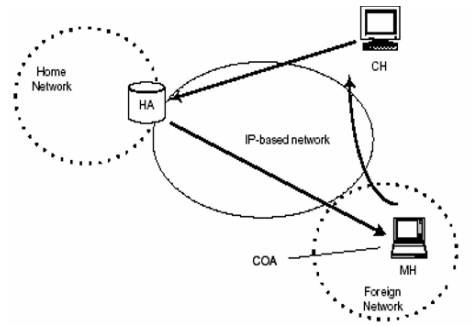


Fig. 3. Mobile IP with collocated COA

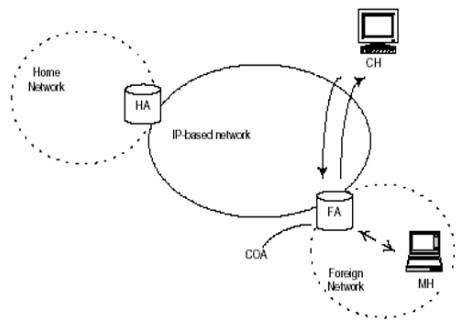


Fig. 4. Mobile IP with Route Optimization

III. MOBILE IP OPTIMIZATION

To solve the problem of triangular routing, MIP with Route Optimization [3] (MIP-RO) has been proposed in Fig. 4. In order to use MIP-RO, a CH must understand binding updates and be able to tunnel packets to a COA, while the MH must send binding updates to the CH to update it on the MH's location. The binding update informs the CH of the COA of the MH and hence the CH can tunnel packets to the COA without going through the HA. Several new messages, including "binding warning", "binding update", "binding request", and "binding acknowledge", are used to maintain the correct COA binding. While MIP-RO deals with the triangular routing problem, it does not address the issue of micro-mobility management. IPv6 is defined in the IETF working group of IP Next Generation [4], by providing enhancements over the capabilities of existing IPv4 service. Basic improvements to IPv4 include optimal header format, reasonable addressing architecture, neighbor discovery mechanism, stateless auto-configuration, and security and QoS support. Mobility support in IPv6 takes full advantage of these enhancements. Three advantages of MIPv6 are apparent: (a) route optimization is facilitated,

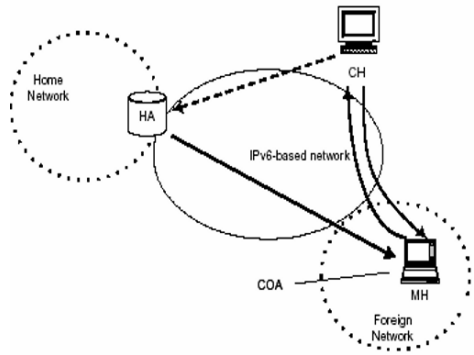


Fig. 5 Mobile IPV6

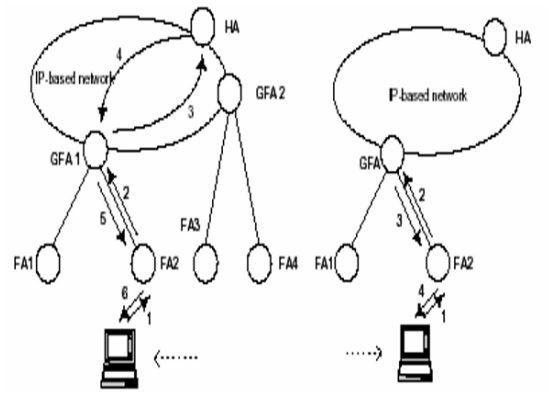


Fig. 6 MIP Regional Registration (a) movement between regions; (b) movement within a region

without needing to be concerned about whether the CHs can understand binding updates, as with MIP-RO; (b) explicit binding updates or MIP registration messages become unnecessary, as the destination options are naturally piggy-backed on IP data packets; and (c) packets from CH to MH need not be encapsulated but are sent directly to the MH with its COA in the source route. The 3rd advantage just mentioned also is due to the way IPv6 makes source routing possible which is shown in fig. 5.

IV. Intra-Domain Mobility Management

To implement a fast and seamless handoff and a minimized control traffic micro mobility management solution are used for intra domain mobility management**.** Micro-mobility solutions that use a hierarchy of mobility agents include MIP with Regional Registration [5] (MIP-RR), and TeleMIP/Intra-Domain Mobility management Protocol [6] (IDMP). MIP-RR perhaps involves the fewest modifications to MIP. In a foreign network, the two level mobility hierarchy contains the upper-layer GFA (Gateway Foreign Agent) and several lower-layer RFAs (Regional Foreign Agent). All MHs under the GFA share the same COA. When a MH moves to another FA under the same GFA, it only needs to register with the new RFA and with the GFA. This is because its HA already knows how to route packets addressed to the MH to that GFA. It does not need to register with its HA unless it moves under a new GFA. Suppose an MH moves between subnets under a GFA with which it is already registered. As shown in Fig. 6(b), the MH initiates its registration with FA2. Then the registration request is sent to GFA1. Since MN is already registered with GFA1, GFA1 does not initiate a home registration to HA, but just sends the registration reply to the MH through FA2. Since the HA does not need to be contacted in this scenario, MIP-RR reduces the handoff latency caused by registration with Home Agent. If the MH changes its GFA, it needs to register with its HA. As shown in Fig. 6(a), the MH moves from FA3 to FA2, and its GFA is no longer GFA2. The MH sends a registration request to its new RFA, which is FA2, and then GFA1. Because GFA1 is a new GFA, it has to register with the HA.

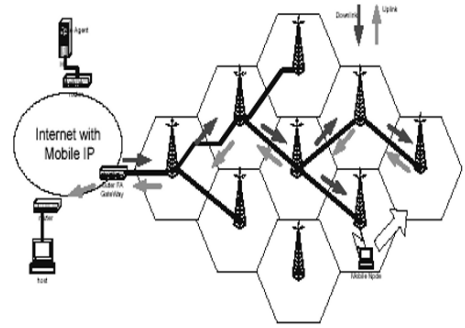


Fig. 7 Cellular IP

V. Cellular Ip

The cellular IP architecture in Fig. 7 consists of interconnected Cellular IP nodes (also known as Cellular IP base stations), these Cellular IP nodes communicate with Cellular IP mobile hosts. The Cellular IP gateway is a special Cellular IP Node as one of its interfaces is connected to a fixed standard IP network. The cellular IP base stations emits beacons on regular basis, this allows the mobile host to locate their nearest base station. When a Mobile host finds its nearest base station it sends a

route update message to its connected base station, this message is routed internally in the cellular IP network from the base stations to Cellular IP Gateway by using a hop-by hop shortest path routing mechanism Cellular IP support two types of handover – hard and semi soft [7]. Hard handoff is usually described as a 'break before make' handoff, i.e. radio link communication is broken before the new radio connection is established. Semi soft uses the fact that certain radio technologies allow nodes to 'listen' to two or more base stations while connected/communicating to one base station. Mobile hosts on a foreign network listen to the base station transmit their beacons and determine if a handover procedure should be initiated by measuring the signal strengths. Once the mobile host receives a stronger beacon from a new base station it tunes itself to the new base station, hence it has stopped listening to the old base station, and sends a routing update message to the new base station. The routing update message creates the uplink\downlink route to the gateway and mobile host is now able to send and receive packets. During most of this time packets are still being sent to the mobile host old base station and hence lost. Since the mappings to the old base station are not cleared by the handover procedure, packets are being sent to both base stations until the mappings time out on the old route. The hard handover procedure is simple but the packet loss is not acceptable for real time applications. Using the fact the Mobile host can listen to the beacons of the base stations and determine when to handover, the Cellular IP protocol uses the Semi soft handover scheme. The Mobile host receives a stronger beacon from a new base station and sends a semi soft packet to it, the mobile host then immediately returns listening to the old base and it's currently connected base station. The semi soft packet sent to the new base station creates the routing cache for the mobile host on the new base station. Once set up data packets are then sent to both base stations, after a semi soft delay time period the mobile host begins the usual handover procedure. This method of handover is better than the hard handover but it uses twice as much of resources than hard handover.

VI. Conclusion

The next generation of wireless communication systems combines both fixed and wireless facilities. We have to ensure uninterrupted delivery of multimedia service as mobile terminals (MTs) move among different areas anywhere in the world. Mobility management plays an important role for the next generation wireless communication and it is necessary to pay special attention to the design of mobility management procedures. In this concern, we have presented Mobile IP as a solution for next generation wireless system networks, which provides secure, macro and micro mobility protocol. Cellular IP works as a micro mobility solution and can work in conjunction with mobile IP for a complete mobility solution

References

[1] G. Fleming et a /, "A Flexible Network Architecture for UMTS," I€€€ Pers ommun, VOI 5, no 2, Apr 1998, pp 8-1 5

[2] G. Fleming et a /, "A Flexible Network Architecture for UMTS," I€€€ Pers ommun, VOI 5, no 2, Apr 1998, pp 8-1 5

[3] C. Perkins, D.B. Johnson, "Route Optimization in Mobile IP", Internet Draft, IETF, November 2000. Work in progress.

[4] IETF IP Next Generation Working Group (ipngwg), http://www.ietf.org/html.charters/ipngwg-charter.html. Progress, July 2000.

[5] E. Gustafsson, A. Jonsson, C. Perkins "Mobile IP Regional Registration", Internet Draft, IETF, March 2001. Work in Progress.

[6] S. Das, A Misra, P. Agrawal, and S.K. Das, "TeleMIP: Telecommunications Enhanced Mobile IP Architecture for Fast Intra-domain Mobility," IEEE Personal Communications Magazine, Vol 7, pp. 50-58, Aug. 2000

[7]M. Carli, "Mobile IP and Cellular IP integration for Inter Access Network handoff" 2001

Addressing Spam at the Systems-Level Through a Peered Overlay Network-Based Approach

Michael Horie, Stephen W. Neville, *Member, IEEE*

***Abstract*—Reducing email spam has been an active industry and academic research domain for a number of years. Despite this, spam has remained an on-going world-wide problem which absorbs significant network resources in its delivery. Client-side solutions have addressed much of the end-user nuisance factor, but trace-back solutions have not succeeded in sufficiently reducing spam ingress at its source due to both the movement towards distributed spam generation and geopolitical factors. At a systems-level, part of the inherent issue in addressing global spam is the current divergence between responsibility and accountability; end-users' are made responsible for addressing issues which the originating ISP's are better positioned to solve. Within this work, an overlay network-based approach is developed, which employs peer-to-peer QoS agreements in conjunction with a non-repudiation protocol for broadcast environments, to affect a low-spam overlay network. This of course does not solve the global spam issue, but does allow participating communities to move to a low-spam environment provided they are willing to accept their agreed to responsibilities.**

I. INTRODUCTION

Spam has remained a core problem for email services world-wide despite the significant efforts of both the academic and industry research communities. Recent studies [1][2] have shown that spam can account for up to 90% of all monthly emails. Such figures translate into upwards of 183 billion emails per day [3] which are propagated through the Internet just to be, ideally, deleted upon receipt. Obviously, the advances in and wide-scale deployment of client and email server-based anti-spam technologies, (*i.e.,* Bayesian filtering, Spam Assassin ™, *etc.*) have significantly reduced the impact of spam for the end-user. The issue remains, though, that large amounts of network resources are consumed to transmit high volumes of spam.

Trace-back approaches are the main mechanisms that have been proposed to address this core issues. Unfortunately, the global nature of the Internet requires that trace-back solutions function within complex geopolitical landscapes, which greatly limits their real-world effectiveness. Additionally, there has been a general move by spammer to the use of large-scale bot networks, which tends to reduce individual machine spam generation to levels below the trace-back thresholds.

Overall, solving the global spam problem requires solving intrinsically hard engineering challenges. In particular, spam can easily be made to look closer to "normal'' email, ultimately requiring a solution to the open problem of machine understanding of context within free-form text. Spam routing can cross and re-cross national boundaries, requiring effective international agreements and levels of trust to allow real-time trace-backs over national IT networks. Some solutions have also been predicated on a re-design of the Internet as a whole, but this is most likely infeasible based on the underlying cost and inertia issues.

From the authors' point of view a real-world effective solution to spam must, for primarily pragmatic reasons, adhere to the following seven observations:

- Abrupt Internet-level technological sea changes tend not to occur, as indicted by history.
- "Opt-in" strategies can lead to sea changes, (*i.e.,* peer-to-peer networks, the world-wide web, MP3 music formats, *etc.*).
- Ultimately, human end-users are the best "detectors" of spam and should be kept "within the loop".
- Spammers will continually adapt to technological changes so solutions must be robust to such adaptations.
- Network resource usage for spam transmission is today's major spam problem.
- Trace-back methodologies must be cognizant of and take into account real-world geopolitical limitations.
- Accountability and responsibility should not be allowed to diverge; the party in the best position to address the given spam issues should be made accountable for addressing it.

Within this work an opt-in low-spam overlay network approach is developed to provide a systems-level mechanism to address spam. This overlay network is inclusive of existing spam reduction techniques and technologies and not intended as a replacement for them. Core to this overlay network are the inclusion of peer-level, locally negotiated, quality of service contracts and a novel non-repudiation protocol suitable for spam's broadcast medium.

The work is structured as follows. Section II will overview the related works available in both the literature and industry to address spam. Section III will then overview the proposed overlay network-based solution. Section IV will then provide the details of both the required peer-level QoS agreements and the non-repudiation protocol, which together enforce the low-spam nature of the overlay. Section V will then verify the protocol via its analysis within the AVISPA security protocol analysis tool. Section VI will then offer areas of future work and conclude the work.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 448–453.

II. Related Works

A significant body of research exists in the area of detecting spam at the leaf nodes. This includes the use of naive Bayesian algorithms [4], support vector machines [5], memory-based classifiers [6], and boosting trees [7]. Commercial products also tend to focus on leaf-node detection, often combining this with complementary techniques such as blacklisting IPs of machines that have sent spam in the past. For example, Cisco's Ironport Antispam [8] evaluates messages based on four characteristics: (a) the reputation of the sender, derived from properties such as normal sending volumes and country of origin, (b) the reputation of included URL links, derived from the age of the domain registration and other characteristics, (c) the structure of the message, including a check for contradictory From: and Received: headers, and (d) the contents of the message. Ironport also operates an operations centre that generates signatures for spam messages that make it passed these four checks.

Cloudmark [9], which grew out of the Vipul's Razor open-source project [10], takes a different approach to detecting spam by using feedback from a large user community. In essence, the program computes a fingerprint for every incoming message and compares this to existing spam fingerprints in a catalogue server. If there is no match, the message is delivered to the recipient's mailbox. In case the recipient feels that the message is spam, the program can be directed to nominate the message as spam. If a sufficient number of users have designated this message as spam, it is forwarded for inclusion in the catalogue server database.

Research has also been conducted in the area of identifying the source of a spam message. For example, Domain Keys Identified Mail (DKIM) [11] uses signatures to identify the sender of a message. Before sending out a mail message, a DKIM-compliant mail server creates a signature for the outgoing mail message, and inserts this signature into the message. The receiving mail server can then query the sending mail server for its public key, and use this key to verify the signature. A mismatch would indicate that a part of the message, such as the From: header, was either modified in transit or forged. Thus messages without DKIM signature, or a mismatched signature, can be separated for further scrutiny. While this is no direct defense against spam, it does allow positive trace-back to the spammer's mail server, exposing it to blacklist inclusion.

A complementary standard, the Sender Policy Framework [12], requires domains to publish the IP addresses of their mail servers. When a message arrives at an SPF-compliant mail server, it checks the given return path of the message. Based on this information, the mail server looks up the DNS record of the domain given by the return path, to get a list of authorized mail servers of that domain. If the message was not delivered by one of these mail servers, it could indicate a forged return path, which is indicative of spam.

SpamCop [13] uses mail headers to make an educated guess about the ISP responsible for injecting a spam message into the Internet. It also allows users to send an automated complaint to the originating ISP, and publishes a blacklist of spamming ISPs as an enforcement mechanism.

Digital Postmarks [14] provide another means to trace offending ISPs. Using this protocol, the first border router along a packet's path inserts a postmark based on the router's IP address, allowing the recipient to narrow down the source of the packet.

In addition to leaf-node detection and trace-back techniques, there have also been proposals to detect spam at the router level [15], to make it economically unfeasible to transmit spam by charging for messages [16][17] or by requiring the sender to compute a non-trivial function [18], or to label e-mail messages according to a universal labeling scheme [19]. Of course, existing techniques, such as [20], which provide mechanisms to trace Denial-of-Service attacks back to their source(s) can also be adapted to provide frameworks for the trace-back of spam.

III. Proposal Overview

The core idea of this work lies in the establishment of a low-spam network overlay over existing commodity Internet, enforced by peer-to-peer QoS agreements. Each participant agrees to take active measures to prevent machines in their domain from injecting spam into this overlay, and accepts responsibility in case the measures are insufficient. In return, each participant gains access to an overlay network in which all other players have agreed to a similar level of responsibility, allowing for a reduction of resource demands due to decreased spam volumes. The measures a participant may chose to enact include such standard methodologies such as prohibiting egress TCP connections to port 25 from unauthorized machines, monitoring egress mailing patterns, and scanning machines in the subnet for bots.

The choice of which measures to enact is left to each participant's discretion; it is expressly not enforced by the proposed protocol. Instead, the protocol is results based. The enacted measures must be sufficient to meet the participant peer-to-peer QoS agreements. Failure to take sufficient active measures would be a violation of these agreements, and depending on the agreement, could result in reduced or denied access to the low-spam network overlay and/or financial penalties.

One of the incentives for backbone providers to participate in the low-spam overlay is the opportunity to sell access to this network as a premium service to lower-level ISPs, corporate organizations, public sector entities, and so on. Those that choose not to pay the premium can continue to use the commodity Internet, while those that do choose the premium service can reduce their exposure to spam, and its use as a malware attack vector, as a mechanism for confidence schemes, and as a potential conduit for information loss. Leaf networks, i.e., lower-level ISPs and corporate and public

networks, would see lower bandwidth requirements, lower server resource demands, and lower disk space needs due to much lower email volumes. For larger companies, these could be reductions of up to 95% [21]. Email tracking requirements to support legislative requirements such as Sarbanes-Oxley, would also be reduced in direct proportion to email volume reductions.

For the end users within these leaf-network, (*e.g.*, ISP subscribers), the main gains would be a reduction in spam, reduction in lost email due to false positives, generally estimated at 5-10% of all valid email sent, and reduced exposure to email based malware, confidence, and Ponzi schemes.

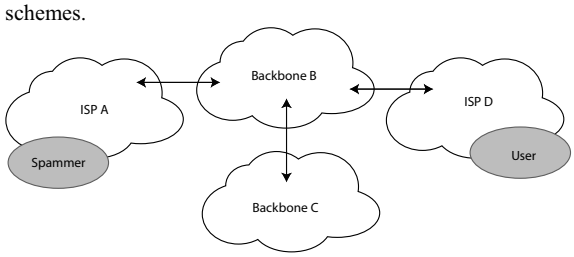


Fig. 1. Interrelation of a spammer and a leaf-node user across their intermediate ISPes and backbone Internet service providers.

The basic idea is illustrated in Figure 1. In this particular example, the user has just received a spam message, and reports it to their ISP, in this case ISP *D*. *D* determines that the spam was sent via backbone *B*. It is important to note that ISP *D* is not required to go after ISP *A* directly. Instead, *D* passes the information regarding spam back up to backbone *B*. As per the QoS agreement between *D* and *B*, *B* must take the next step.

If *B* fails to do so then the penalty terms of this agreement come into play and *D* gains the agreed upon compensation. Likewise, backbone *B* passes the issue to ISP *A* with which it has another peer-to-peer QoS agreement. *A* must then address *D*'s identified spam issues or face the penalties under its agreement with *B*. If *A* addresses the problem as required, then *B* meets its terms with *D* and the problem is solved.

If not then *B* can choose to either accept compensation from *A* which can be used to mitigate the cost of the compensation due to *D*, or in the extreme case, choose to sever *A* from the overlay, and in that manner address *D*'s concern. This approach allows spam trace-back to be implemented in a manner which is cognizant of geopolitical realities. The penalties that peers can place on each other are based on their QoS agreement, which they both freely entered into. Such agreements offer a significantly-reduced level of legal complexity over agreements executed at the nation-to-nation level. Hence, the action available to *B* against *A* can be enacted within the timeframes expressed by the QoS agreement, which may be immediate, if this has been authorized in the QoS terms.

IV. SPAM TRACING PROTOCOL DETAILS

A. Sender to Recipient Path

Every network N_i that is part of the overlay marks SMTP packets at the network's ingress points with the identity of the router in neighboring network N_{i-1} that was responsible for injecting the packet. The next network down the path, N_{i+1}, similarly marks the packet with the identity of the egress router of N_i. This is repeated until the packet is delivered to an MTA. When spam filters or users subsequently report an e-mail to be spam, this information allows the participants to determine who violated the QoS agreement, and take appropriate action.

More specifically, an ISP that has subscribed to the low-spam network overlay must ensure that all its mail servers sign every outgoing mail message that is to be transmitted over the low-spam network. In addition, when the mail message is broken into separate IP packets, every packet will have to be signed as well. Backbones will need access to this signature, so this signature must be transmitted at the IP level. Spam Tracing follows protocols like IPsec [22] and MPLS [23], by implementing a bump-in-the-stack approach: When first being placed onto the overlay, the router must inject an IP packet whose headers are identical to the original packet, except that the protocol type is now set to the Spam Tracing protocol. The payload of the new packet initially carries the number of label pairs, the result of signing the e-mail signature and the original packet with the ISP's private key, the e-mail signature, and the original IP packet. This is illustrated in Figure 2.

Label Pairs =1	A's Signature	E-mail Signature	Original IP Packet

Fig. 2. Packet generated by ISP A and transmitted to Backbone B.

The neighboring network operates two types of routers. Routers connected to a network that did not sign up for the low-spam service, route packets normally. Depending on existing agreements, there may be an option to treat IP packets from such networks as low-priority bulk packets, as long as packets have a TCP payload headed to port 25. Routers connected to a network that did sign up for the low-spam service, proceed as follows: If the ingress router detects an IP packet that contains a TCP payload headed to destination port 25, the router assumes that it is dealing with an SMTP connection, and will execute the Spam Tracing protocol. Otherwise, the packet will be routed normally.

Label Pairs =2	---	A's Egrees Router IP	A's Signature	E-mail Signature	Original IP Packet

Fig. 3. Packet generated by the ingress router of Backbone B.

The ingress router verifies ISP A's signature against the e-mail signature and original IP packet. If it does not match, the packet is discarded. If it does match, the number of label pairs

field is incremented by 1. Then the IP address of the neighboring router from which the packet is arriving, and a placeholder for the backbone signature, is placed in front of the ISP's signature fields. This is illustrated in Figure 3.

The IP packet is then routed over the low-spam overlay. The egress router of backbone B inspects the packet. If the protocol is the Spam Tracing protocol, and the adjacent network does not understand the Spam Tracing protocol, the original IP packet is restored. This removes the ability of downstream networks to lodge a complaint based on the Spam Tracing protocol; however, since they did not sign an appropriate peer-to-peer agreement, this is a natural consequence. The email is back to commodity Internet transmission under no low-spam QoS agreement(s). If the neighboring network did sign up for the overlay, the backbone replaces the incoming neighboring router IP with a lookup key. There is a unique lookup key for every (incoming neighboring router IP, outgoing neighboring router IP) pair. This lookup key and the ISP A fields are then signed by backbone B with its private key. This is shown in Figure 4. Depending on the level of security desired by B, it may choose to use one (private key, public key) pair for every neighboring network, or use only one key pair for all networks.

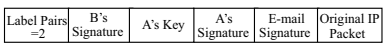


Fig. 4. Packet generated by the egress router of Backbone B.

The next network verifies backbone B's signature against the other fields. If they do not match, the packet is discarded. Otherwise, the next network increments the number of label pairs by 1, and adds backbone B's egress router's IP address and a placeholder for the current network's signature. Just before reaching the recipient MTA, the IP address is replaced with a lookup key, and the packet is signed, as shown in Figure 5.

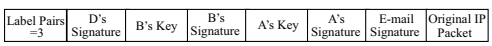


Fig. 5. Packet generated by ISP D, immediately prior to the packet being handed off to the User.

The machine hosting the MTA stores the packets associated with the current SMTP connection against a unique lookup key. A Spam Tracing protocol-aware MTA must inject this key as an SMTP header into the mail message. To deal with legacy MTAs, one can construct a lightweight helper application listening on port 25, whose only responsibilities are to capture and store the IP packets associated with an email connection, forward the SMTP commands to the legacy MTA listening on another port, and to inject the spam key header at the appropriate point. It must also make sure that any e-mail whose signature does not match the e-mail signature in the IP header, is dropped.

Users can then report spam by submitting the spam key header(s), which will pinpoint the neighboring network. These reports can then be used to pinpoint which neighboring networks were responsible for a particular spam message.

B. Trace-back

Detection of spam may be via any of the available leaf-node detection approaches, including the human user marking the email message as spam via their email application. This detection is then reported to ISP *D* by means of submitting the original IP packets to *D*. ISP *D* checks the e-mail against ISP *A*'s e-mail signature and the spam trace packets against ISP *D*'s signature. Should either one of them fail, the e-mail message was tampered with, and is rejected. The tampering party's machine can then be identified and appropriate follow up actions be taken. For example, the tampering machine may be a compromised bot network machine which needs to be cleaned, or it may be an intentional tampering by the machine's owner, in which case the ISP-to-end user agreement's terms come into play to address such activities.

If the signature checks pass, ISP D uses the lookup key to determine that the next step in the upstream chain was Backbone B. Depending on the peer-to-peer agreement, the complaint is either directly forwarded to B, or on composite only after a certain threshold of complaints against B have accumulated. The spam header that is forwarded to B is truncated, with ISP D's signature and Backbone B's lookup key removed. If the threshold of complaints approach is employed, then it is assumed that all emails associated with the threshold being exceeded are packaged in bulk by D and sent upstream to B.

Backbone B receives the e-mail and spam headers, and checks the e-mail signature against the e-mail, and the spam packets against Backbone B's signature. Again, should either of them fail, the message can be considered to have been tampered with, and can be rejected. In this case, Backbone B knows that ISP D was the party from which the tampering arose and can take the appropriate actions as detailed within their QoS agreement.

If the signature checks pass, Backbone B uses the lookup key to determine that ISP A injected the spam. Backbone B can now lodge a complaint with A.Should ISP A fail to act, Backbone *B* can deny ISP *A* future access to the low-spam overlay.

C. Overhead

The overall packet size grows linearly with respect to the number of backbones that the packet traverses. Specifically, the source network adds a label count and two signature fields to the original IP packet. Given that an IP packet cannot live past 256 hops, the label count requires 8 bits at the most. The protocol does not specify which signature algorithm to use, but assuming that a reasonably-efficient short signature algorithm like the one proposed by [24] were to be used, each signature field would require 119 bits per signature, for a total of 8 + 2 * 119 = 246 bits. Every backbone, as well as the

destination, then adds a lookup key and a signature. The lookup key is used to uniquely identify an ingress network. Given that the Autonomous System (AS) Number is 32 bits long, this would add 32 + 119 = 151 bits to the packet per hop. With a common AS path length of about 5 [25], this yields 246 + (5 * 119) = 841 bits, or 106 bytes. The destination mail servers have to store these packets for a certain period of time, to be able to facilitate a complaint under the peer-to-peer protocol. Given a TCP/IP header overhead of 40 bytes, this means that at most 146 bytes per packet will have to be retained. Storage concerns can be addressed by requiring a response within a certain time frame, such as 7 days after receipt of the spam message.

V. FORMAL VERIFICATION

The Spam Tracing protocol is essentially about non-repudiation of origin: It must not be possible for the originator or conveyor of a packet to deny involvement. Formal verification of this property has proved to be a challenge, in part because the body of research in this area is much smaller than in the areas of verifying secrecy and authentication [26]. AVISPA [27] is a tool that has been used to verify security and authentication properties of many different protocols. Recently, this tool was adapted to allow verification of non-repudiation properties of protocols such as Fair Zhou-Gollman [28].

For verification of non-repudiation, AVISPA provides two predicates, *witness* and *wrequest*. *witness(A,B,x,Y)* means that an agent *A* has communicated message *Y* to agent *B*. *wrequest(B, A, x, Y)* means that agent *B* has received a message *Y* which it believes to be from agent *A*. *x* is simply an identifying label for this statement, to make it easier to understand attack messages. To ensure non-repudiation, there must never be a *wrequest* without a prior matching *witness*. Otherwise, the protocol in question allows an intruder to impersonate the originator.

An issue arises when trying to use these predicates in a broadcast scenario such as spam. For example, consider A → B: M.$\{H(M)\}_{inv(Ka)}$, where agent *A* sends a message *M* to agent *B*, along with a hash of *M*, signed by *A*'s private key. To implement this in AVISPA, A must issue the statement *witness(A, B, a_to_b, $\{H(M)\}_{inv(Ka)}$)* and *B* must issue the statement *wrequest(B, A, a_to_b, $\{H(M)\}_{inv(Ka)}$)* to indicate that the signature can only have originated from *A*. However, the AVISPA tool will raise an attack alert, because it is possible that *B* will receive and accept a message from *A* that was not meant for *B*. More specifically, if *A* sends a message intended for an agent other than *B*, *witness(A,B,x,M)* will not be raised. However, an intruder is able to copy *M* to *B*, causing *B* to raise a *wrequest(B,A,x,M)*. This means that there is a *wrequest* without a prior matching *witness*, and so AVISPA raises an attack alert.

To circumvent this tool issue, one can insert a virtual trusted third party (TTP). The message exchange between two Spam Tracing peers can then be modeled as follows:

1. A → B: $M_n.M_{sig}.In_{sig}.In.Out_{sig}$
2. A → TTP: $\{A.B.Out_{sig}.Cnf_{sigA}\}_{KTtp}$
3. B → TTP: $\{A.B.Out_{sig}.Cnf_{sigB}\}_{KTtp}$

where:

- M_n is a fragment of the overall message,
- M_{sig} is the signature $\{H(M_1...M_n.)\}_{inv(KOriginator)}$,
- In_{sig} is the signature $\{H(M_n.M_{sig})\}_{inv(KIn)}$,
- In is the identifier for the incoming neighbor,
- Out_{sig} is the signature $\{H(M_n.M_{sig}.In_{sig}.Agent)\}_{inv(KA)}$,
- Cnf_{sigA} is the signature $\{H(A.B.Out_{sig})\}_{nv(KA)}$, and
- Cnf_{sigB} is the signature $\{H(A.B.Out_{sig})\}_{nv(KB)}$

The *witness* predicate is raised by *A* after step 1. Steps 2 and 3 are part of the virtual TTP, which was added to address the previously-discussed AVISPA issue. To ensure that these do not influence the protocol model, the TTP is a passive observer whose only purpose is to raise a matching *wrequest* predicate after hearing from A and B. After running this model, AVISPA did not find any attacks, which though not a conclusive proof, raises confidence in the protocol.

VI. CONCLUSIONS AND FUTURE WORK

This work has presented a systems-level approach to addressing spam through the development of opt-in low-spam overlay networks. The low-spam nature of these overlays is jointly supported and enforced through peer-to-peer QoS agreements and the developed non-repudiation protocol, which is specifically structured to support the innate broadcast nature of spam. Taken together these provide spam recipients with mechanisms to identify and force upstream conveyors of spam to take action against its transiting of the systems they control, as per the legal requirements embodied in their signed peer-to-peer QoS agreements. Verification of the core elements of this approach and protocol were validated through the AVISPA tool. The novelty of this approach lies in its joint use of peer-to-peer QoS agreements and the non-repudiation protocol to affect a systems-level approach to spam in which accountability and responsibility are inherently directed to those in the best position in the overlay to counter each given instance of spam ingress.

The larger-scale verification of a complete end-to-end session is ongoing. Such end-to-end sessions, especially when inclusive of multi-packet emails and IP packet fragmentation, have significantly larger search spaces which AVISPA must consider. The size of these search spaces is proving to be a challenge under the current version of the AVISPA tool which the authors are employing.

Obviously, the approach presented is not structured to provide an instantaneous solution or response to each given instance of spam. Instead the peer-to-peer QoS agreements enforce limits on aggregate spam levels and, hence, an inherent time delay is required to assess aggregate levels. Additionally, it can be observed that the benefits of the overlay only accrue to the overlay participants, (i.e., those who have chosen to opt-in to peer-to-peer QoS agreements).

Hence, it can be argued that, as with most network-based technologies, the value of the network depends on the number of network "members", leading to the obvious "chicken and egg" problem at startup. Of course, many existing Internet technologies have suffered similar startup concerns. The World Wide Web is of limited value if only a small number of CERN researchers are using it. Such technologies have gained prominence in part due to the combination of their usefulness and their support of an opt-in adoption philosophy, which the proposed solution has been structured to support.

Such observations though do lead to important questions regarding: (a) the expected adoption rate of the proposed technology, (b) the effect of the adoption rate on the solution's efficacy, (c) the potential dependency of the QoS agreements' feasible spam tolerance thresholds on both time and adoption rates, (d) whether the existence of sufficient large overlays would "push" spam away from the overlay and into generic (bulk) email propagation paths, (e) whether differential billing for access to low-spam overlays is a sufficient incentive to accept the responsibilities enforced by the QoS agreements, and whether (f) leaf node networks will have a sufficient incentive to opt into this protocol. Answering these, and a number of other relevant questions, is the subject of on-going work. More specifically, a simulation-based approach along with an appropriate game theory framework is being developed to more formally explore these questions. Until the simulation data are available, these questions remain open.

VII. Acknowledgements

The authors would like to thank Prof. Laurent Vigneron for his kind assistance in their use and understanding of the AVISPA tool. All issues and challenges in the use of AVISPA remain solely with the authors.

VIII. References

[1] http://www.messagelabs.com/publishedcontent/publish/threat_watch_dotcom_en/intelligence_reports/february_2007/DA_181260.chp.html (Current as of October 11, 2007)

[2] http://www.postini.com/stats/ (Current as of October 11, 2007)

[3] http://www.businessweek.com/technology/content/sep2006/tc20060919_412904.htm (Current as of October 11, 2007)

[4] P. Pantel and D. Lin, "Spamcop: a spam classification & organization program," Proc. AAAI-98 Workshop on Learning for Text Categorization, pp. 95-98, 1998.

[5] H. Drucker, W. Donghui, and V.N. Vapnik, "Support vector Machines for spam categorization," IEEE. Trans. on Neural Networks, vol. 10, no. 5, pp. 1048-1054, September 1999.

[6] I. Androutsopoulos, G. Paliouras, V. Karkaletsis, G. Sakkis, C.D. Spyropoulos, and P. Stamatopoulos, "Learning to Filter Spam E-Mail: A comparison of a naive bayesian and a memory-based approach," Proc. Machine Learning and Textual Information Access, 4th European Conference on Principles and Practice of Knowledge Discovery in Databases, pp. 1-13, September 2000.

[7] X. Carreras and L. Marquez, "Boosting trees for anti-spam email filtering," Proc. Recent Advances in Natural Language Processing, pp. 58-64, 2001.

[8] http://www.ironport.com/pdf/ironport_anti-spam_whitepaper.pdf (Current as of October 11, 2007.

[9] http://www.cloudmark.com/releases/docs/wp_reputation_filtration_1064 0406.pdf (Current as of October 11, 2007)

[10] http://razor.sourceforge.net/ (Current as of October 11, 2007)

[11] http://www.ietf.org/internet-drafts/draft-ietf-dkim-overview-05.txt (Current as of October 11, 2007)

[12] http://www.openspf.org/Introduction (Current as of October 11, 2007)

[13] http://www.spamcop.net/ (Current as of October 11, 2007)

[14] T. Costlow, "Digital postmark helps fight spam, virus attacks," IEEE Distributed Systems Online, vol. 5, no. 5, 2004.

[15] B. Agrawal, N. Kumar, and M. Molle, "Controlling spam emails at the routers," Proc. IEEE International Conference on Communications, vol. 3, pp. 1588-1592, May 2005.

[16] R. L. Rivest and A. Shamir, "Payword and micromint: two simple micropayment schemes," Proc. 4th International Security Protocols Conference, pp. 69-87, 1996.

[17] D.A. Turner and D.M. Havey, "Controlling spam through lightweight currency," Proc. Hawaii International Conference on Computer Sciences, January 2004.

[18] C. Dwork and M. Naor, "Pricing via Processing or Combatting Junk Mail," Lecture Notes in Computer Science, vol. 740, pp. 139-147,1992.

[19] A. Herzberg, "Controlling spam by secure internet content selection," Proc. 4th International Conference on Security in Communication Networks, Lecture Notes in Computer Science, vol. 3352, pp. 337-350, 2004.

[20] S. Savage, D. Wetherall, A. Karlin, and T. Anderson, "Network support for ip traceback," IEEE/ACM Transactions on Networking, vol 9, no. 3, pp. 226-237, June 2001.

[21] http://radar.oreilly.com/archives/2007/03/another_war_wer.html (Current as of October 11, 2007)

[22] S. Kent and R. Atkinson, Security architecture for the internet protocol, RFC 2401, 1998.

[23] E. Rosen, A. Viswanathan, R. Callon, "Multiprotocol label switching architecture, " RFC 3031, 2001.

[24] N. Courtois, M. Finiasz, and N. Sendrier, "How to achieve a McEliece-based digital signature scheme," Proc. Asiacrypt '01, Lecture Notes in Computer Science, vol. 2248, pp. 157–174, 2001.

[25] Y. Harada, K. Okamura, T. Chiyonobu, and Y. Lee, "Analyzing Correlation Between Flow Data and AS Paths in BGP Routing," Proc. Frontiers of High Performance Computing and Networking – ISPA 2006 Workshops, Lecture Notes in Computer Science, vol. 4331, pp. 1126-1135, 2006.

[26] J. Santiago and L. Vigneron, "Optimistic non-repudiation protocol analysis," Proc. 1st Workshop on Information Security Theory and Practices. Smart Cards, Mobile and Ubiquitous Computing Systems, Lecture Notes in Computer Science, vol. 4462, pp. 90-101, 2007.

[27] A. Armando, D. Basin, Y. Boichut, Y. Chevalier, L. Compagna, J. Cuellar, P.H. Drielsma, P.C. Heám, O. Kouchnarenko, J. Mantovani, S. Mödersheim, D. von Oheimb, M. Rusinowitch, J. Santiago, M. Turuani, L. Viganò, L. Vigneron, "The avispa tool for the automated validation of internet security protocols and applications," Proc. Computer Aided Verification, Lecture Notes in Computer Science, vol. 3576, pp. 281-285, 2005.

[28] J. Zhou and D. Gollmann, "A fair non-repudiation protocol," IEEE Symposium on Security and Privacy, pp. 55-61, 1996.

A Step Towards an Autonomous Tuning Engine Design for Self-Protection and Self-Configuration

Nadir Zamin Khan, Andreas Mitschele-Thiel
Faculty of Computer Science and Automation
Technical University Ilmenau, Germany.
and Danish Naeem
LSS – CNRS – Supélec – Université Paris-Sud 11
Plateau de Moulon, 91192 Gif-sur-Yvette, France

Abstract- **The efficiency of any system is related to the quality of its configuration and tuning. Especially in security appliances, there exists strong relativity in performance efficiency and quality of tuning and configuration. This paper presents an approach to design a tuning engine that will dynamically configure the security devices according to the rules defined by policy makers and network administrators. The tuning engine will dynamically configure the devices to act efficiently for specific threats. It will filter the irrelevant daily alerts from different security devices. The scope of response of a security system is selected by finding out the probability of occurrence of a specific threat and the severity of that threat. Thus the tuning engine dynamically configures the response of security system according to the policies defined for different types of threat. This is an attempt to design a tuning engine which has influences from general policies, administrator settings and information during its operations. The mechanism of introducing the influence of information and policies for tuning and configuration of security devices will efficiently decrease the false alarm rates in any security system.**

Index Terms - Autonomic Computing, Security Matrices, Self-Configuration, Self-Protection, Tuning Engine

I. INTRODUCTION

In dynamically changing environments, security sensor products usually generate volumes of alerts, only little of them are valid, the remaining are false alarms. Reporting from devices is not coordinated and there are no correlations with the reports generated on daily bases from different vendor's solutions (i.e. Mobile Sensors, Host-based intrusion detection system-HIDS and Network intrusion detection system-NIDS). The problem of effective reporting, complexity in management and implementations are the key factors related directly to the quality of any security system. Administrators face problems like information overloading, information handling, legitimate threats, decentralized reporting infrastructure, and higher installation, maintenance and configuration time. The quality of daily actions, automated or manual, is very much related to the quality of response. Typically the response time and the effective threat mitigation efforts are important for success of any security system.

An approach to design tuning engines that will dynamically configure the intrusion detection system (IDS) or intrusion prevention system (IPS), including sensors and devices, is addressed in this paper. The configuration and tuning is done according to the rules defined by policy makers and network administrators. The lack of qualitative configuration and tuning of these devices is an issue which leads to less effectiveness of whole system. The relationship between the probability of detection and false positive rate of IDS leads to the fact that IDS can be configured or tuned to favour either the ability to detect attacks or to minimize false positive rate. To autonomically choose the trade off between these two options, one must have to define a way to autonomically select scope and goals of system response. A series of variables reflecting information and response of system for specific threat should be found first to have such a mechanism. The scope of response is set, depending on the effects of these variables. For example, in case of security management of the network, scope of response should be defined according to the time, security needs, standards being followed, costs, skills and threat levels. A tuning engine will use these variables to dynamically tune and configure the devices according to predefined operational goals by policy makers and administrators. The main idea of our approach is to have an autonomic tuning engine to efficiently configure and tune the security system of any network. This mechanism will enhance the efficiency and will reduce some of administrator's daily tasks.

II. RELATED WORK

Generally the engineering communities have been using either feedback control approaches [2], [3] or feed-forward control approaches [6] to solve the problems of autonomic computing. In [6], a general method for constructing predictive autonomically tuned systems was outline, which was called as "Clockwork" method. The main difference with our approach is that they did not define any mechanism of interaction or influence of policy and decision makers in their method. A framework is proposed which combines feedback control approach with feed-forward control approach in [7]. They are working on the relationship between performance metric and performance tuning parameters. The research is more focused on self-configuration of server side components. Their solution is very specific to server side component configuration. In [5], they tried to increase the system performance of a data grid by developing matrices for probabilities of transactions. Decisions are made according to statistic prediction algorithms. In [1], they developed an online multivariate analysis algorithm

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 454–457.

named as MANA which is based on multivariate statistical technique [4] to analyze the behavior of system resources and network protocols in order to proactively detect network attacks. The difference is that they are not defining the mechanism to select the scope of response of the security system. This mechanism of tuning engine design with weights of variables being calculated along with the definition of scope of response matrices can be used to develop other tuning engines towards self-optimization and self-healing. This mechanism will be further used, for the major goal to achieve self-management (i.e. self-configuration, self-optimization, self-protection and self-healing, all in one infrastructure).

III. TUNING ENGINE

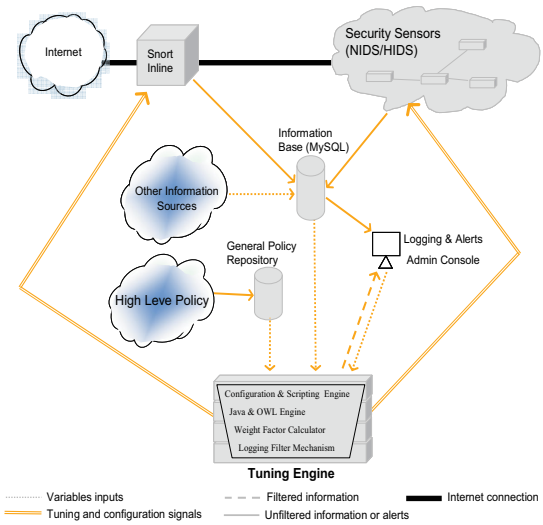


Fig. 1. Tuning engine is shown configuring the security devices and filtering logs & alerts for administrators.

In Fig. 1, a network is connected to the internet while an IPS (i.e. Snort inline) is used in the middle of the connection. This is shown by a thick line connecting the network to the internet. IPS along with some security sensors, in the network, log security related information in the information base. Usually this information can be seen, by administrators, at a console. Information related to different threats can also be found in other information resources, provided by vendors or other security specific communities. Administrators can manually configure the security devices too. In this paper, the tuning engine is introduced to self-configure the security devices.

In Fig. 1, a tuning engine and its interactions with other security devices are shown. The tuning engine has different sub-components such as configuration and script engine, Java and Web Ontology Language (OWL) engine, weight factor calculator, and logging filter mechanism. Configuration and scripting engine sends the configuration instructions to security devices and sensors during tuning operations, as shown by the doubled lines in Fig. 1. The Java and OWL engine uses ontology taxonomies for tuning operations. Weight factor calculator, calculates the weights of different necessary variables. These variables along with other information are used to tune overall operation of the security system of the network. Variables are input from general policy repository, information base and some settings from administrators, as shown by dotted line in Fig. 1. The logging filter mechanism adjusts the relevance of different alerts. Logging filter will sent only relevant alerts to administrators according to defined settings and policy of the network. It is represented by a dashed line in Fig. 1. Unfiltered alerts or information could be stored in an information base for record sake. High level policies are stored in general policy repository.

A. Scope of Response Matrices

The security and vulnerabilities management should have a clear definition of its scope of response. This is a good practice for a cost effective and efficient implementation of security solutions. The first step of our tuning engine is to decide the scope of response. Some matrices are defined that will help in selecting the scope of response for specific threats.

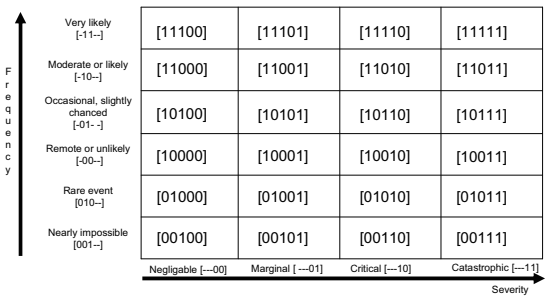


Fig. 2. An example of scope of response matrix.

In Fig. 2, one of these matrices is shown. This matrix is specific for selecting the scope of response of the system particularly for a particular threat. This matrix set the scope of response according to the severity and frequency of a certain threat. The choice of severity of a particular threat on the x-axis is usually input from the information base, depending on the known effects of that threat. The frequency of that specific threat on the y-axis is found by calculating the probability of that threat to occur in the network. The probability is calculated with different variables. Variables are the input from the information base, policy makers and pre-defined administrator settings for a particular network. Chosen values of both axes will decide the scope of response from this matrix. The response details will be clearly defined in the policy repository (i.e. to do what at particular value of matrix). In the next section the mechanism of choosing the frequency of a particular threat by calculating the probability of that threat is shown. The relationship among different variables influences the choice of frequency of a particular threat.

B. Probability calculation and relationship among different variables

Exciting relationships exists among time, security, standards, and costs when considering the probability of any threat to occur in any network. From empirical knowledge of networks and threats, it can be easily said that increasing fame of any system leads to higher threat levels. There is a high probability of attacks as the time of a system being up increases. High risk is related to high probability of being compromised. The probability of being compromise increases while following a well known standard but low cost of implementation is a benefit of that use. Mostly, low cost of implementation in term of time and skills leads to high risk. Following a well-known standard leads to less time of implementation with lesser number of bugs. To calculate the probability of occurrence of any threat, a mechanism is defined.

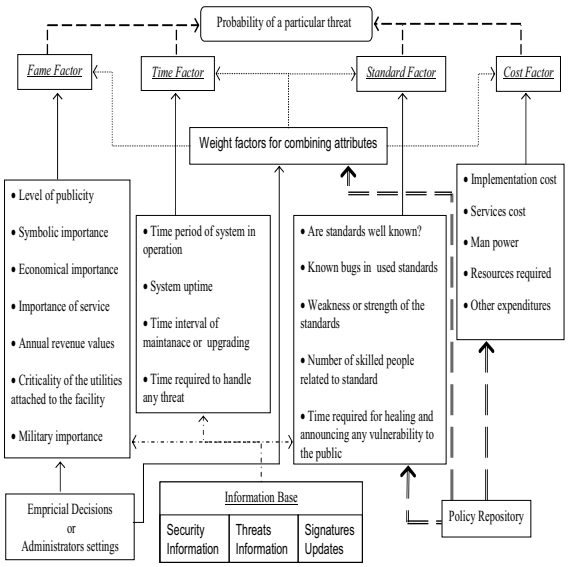


Fig. 3. Calculation of probability with the help of different factors.

In Fig. 3, different factors that will influence the probability are shown. Probability of threat is directly calculated by variables such as fame factor, time factor, cost factor and standard factor. As shown in Fig. 3, each factor will be taken by as a measure of the public estimation impact of the system's operation, computed as a weighted combination of the different attributes of the system. The empirical decision and administrators setting will affect the measure of the public estimation impact of the system's operation for fame factor variable. The output of weighted combination of different factors is getting inputs from policy repository and empirical decisions and administrators settings. The standard factor and cost factor are influenced by policies too. The fame factor, time factor and standard factor is getting some input from information base. The attributes such as military, economical and symbolic importance of system, level of publicity, revenue values, criticality of the utilities attached to the facility affects the fame factor. Similarly the attributes like system uptime, operation time, time interval of maintenance or upgrading of patches, time required to handle any threat influences the time factor. The standard factor is influenced by attributes like number of known bugs in the used standards, weakness or strength of the standard, availability of skilled people related to standards, time required for healing and announcing any vulnerability to the public. The cost factor has attributes like cost of service, staff, resources required for security and other expenditures. Empirical Bayes methods and Dempster-Shafer theory of evidence is used to work on the probability variables. These mathematical theories are currently being worked on to get some workable results. Many other variables like probability of successful attacks or probability of expected attacks can also be found out by the above define mechanism. The tuning engine will dynamically configure and tune the security system according to the scope of response matrices. The severity of a threat is selected with the input from the information base. The frequency of a threat will be chosen by calculating the probability of a particular threat. Three different sources namely, information base, empirical decisions and administrator settings, and general policies makers will influence the probability calculations. The java engine and weight calculators will use some weight factors, probability theories and algorithms to calculate the probabilities.

C. Mathematical expression

Fame Factor (Fa), Time Factor (Tf), Standard Factor (Sf), and Cost Factor (Cf) are considered to have proportional relationship to probability of a particular threat (Popt) for a given system. Popt can be written as follows:

$$\underline{Popt} \quad \sim \quad [\underline{Fa} \,.\, \underline{Tf} \,.\, \underline{Sf} \,.\, \underline{Cf} \,.\, \underline{\phi}] \qquad (1)$$

$\sim$ is representing the proportional relationship. $\underline{\phi}$ can be taken as a measure of any factor affecting the security of a specific system.

Each of the factors in (1) can be assigned accordingly. For example, simple approach could be to assign the number between 0 and 1, computed using a multi-variate utility method. This method computes each factor as a summation of pre-defined weighted values. These weighted values can be between 0 and 1 for the attributes that are defining a certain factor. Here Fa, Tf, Sf and Cf are those factors.

$$\underline{Fa} = \Sigma\,[wt_k * y_k(x_k)] \qquad (2)$$

$$\underline{Tf} = \Sigma\,[wt_k * y_k(x_k)] \qquad (3)$$

$$\underline{Sf} = \Sigma\,[wt_k * y_k(x_k)] \qquad (4)$$

$$\underline{Cf} = \Sigma [wt_k * y_k(x_k)] \quad (5)$$

x_k can be any value of attribute k (e.g., very low). $y_k(x_k)$ is a function that relates x_k to a value between 0 and 1, let us say, for example, very low corresponds to 0. wt_k is a weight factor on k. Σ denotes summation of all attributes for a certain factor.

Pre-defined matrices and variables are used for selecting the scope of response, tuning strategies and control mechanisms. Empirical Bayes methods and Dempster-Shafer theory of evidence are used to predict different probabilities. Other variables such as Probability of successful attack attempts *(Posa)* & Expected attacks *(Ea)* can be depicted by using these mathematical theories along with (1). This tuning engine uses Java and OWL languages to be able to self-protect and self-configure. Same mechanism of using weighted factor can be used to design different engines towards self-management.

IV. CONCLUSIONS AND FUTURE WORK

The tuning engine design makes use of predefined matrices and variables to select the scope of response, the tuning strategies and control mechanisms. The tuning engine mechanism is under development in TU-Ilmenau. Java and OWL languages are used. Mathematical theories such as Empirical Bayes methods and Dempster-Shafer theory of evidence are being used to predict the probabilities. The tuning engine is a step towards wider designing and development of other engines towards self-management in our network.

A next step would be to test the performance of this tuning engine. More clear definition of ontology taxonomy, matrices and variables will be made available. Claims of relationships that exist between the defined variables for a security system will be tested. After successful testing of this engine, the same mechanism will be used to develop other engines towards self-management.

REFERENCES

[1] Multivariate statistical analysis for network attacks detection, 2005.
[2] T.F. Abdelzaher and N. Bhatti. Web server qos management by adaptive content delivery. In Quality of Service, 1999. IWQoS '99. 1999 Seventh International Workshop on, pages 216–225, 31 May-4 June 1999.
[3] Baochun Li and K. Nahrstedt. A control-based middleware framework for quality-of-service adaptations. IEEE Journal on Selected Areas in Communications, 17(9):1632–1650, September 1999.
[4] Douglas C. Montgomery. Design and analysis of experiments 5th edition. New York: John Wiley., 2001.
[5] María S. Pérez, Alberto Sánchez, Ramiro Aparicio, Pilar Herrero, and Manuel Salvadores. Providing autonomic features to a data grid. Pages 337–346, 2005.
[6] L. W. Russell, S. P. Morgan, and E. G. Chron. A new movement in autonomic systems. IBM Systems Journal 42, 1, 2003.
[7] Yan Zhang, Anna Liu, and Wei Qu. Software architecture design of an autonomic system.

Enhancing Network Performance with TCP Configuration

Napat Sra-ium and Kobchai Dejhan

Faculty of Engineering and Research Center for Communications and Information Technology
King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok 10520, Thailand
Tel : 66-2326 4238, 66-2326 4242, Fax : 66-2326 4554
E-mail : {napat, kobchai}@telecom.kmitl.ac.th

ABSTRACT

Presently, the INTERNET network has been considerably congested due to the excessive use of the service. Because the current TCP Reno, a protocol suite in a transport layer, doesn't correlate to the use in WWW scenario, the overall performances of a network are subjected to be laborious to amend. In this paper, we propose the technique to improve the performance of the TCP in a congestion condition. By using the modification algorithm based on TCP Vegas, we could showing how control parameters of TCP Vegas should be configure for achieving better both performance and better throughput.

1. INTRODUCTION

As the Internet exponentially expands the traffic management emerges as the major problem of the network. The Internet which is the most world comprehensive network uses the TCP/IP as its major protocol. The advantage of this protocol is its hardware independent. Therefore it can be used with any type of the networks regardless of their platforms. Although both TCP and UDP are the two major protocols at the transport layer of the network most of the services and application programs usually use TCP for example WWW using HTTP and FTP, electronic mail using SMTP. It is estimated that 90% of the Internet traffic uses TCP particularly the most popular one, WWW. According to the study of the current Internet status, more than 50% of the traffic of the network is due to WWW. Increasing the number of WWW user will increase the traffic. Consequently the improvement of the TCP performance will definitely increase overall performance of the network.

Theoretically, the protocol of the transport layer should be independent and transparent to the technology of its highers layer in the network. The TCP on the fiber optic network for example should has the same performance as the TCP on the cell phone network. However, in practice different media results in different performance. This is because TCP was originally designed for a wire media. When the packet lost, the algorithm of the sender will decrease the transfer rate. In contrast, the wireless network will resend the packet thus increasing more data. In addition, with several new communication technology such as mobile phone, DSL or cable modem, V.90 analog modem, Fast Ethernet and Gigabit Ethernet, and QoS the improvement of the TCP performance are also different.

Currently TCP widely used in the Internet are either Tahoe or Reno or their derivatives. BIC and CUBIC TCP is used in Linux operating system. TCP Reno is used in Windows and UNIX operating systems. Because of its low efficiency of its throughput and fairness other TCPs have been developed such as TCP Vegas [1] which is 40% more efficient than TCP Reno. However when TCP Vegas and TCP Reno are employed in the heterogeneous network in some cases TCP Reno is more efficient. Therefore, this paper proposes a method to enhance performance of the network in terms of throughput as well as fairness by adjusting certain parameters of TCP Vegas in order to achieve a better performance than TCP Reno.

2.TRANSMISSION CONTROL PROTOCOL

TCP is a connection-oriented protocol of the transport layer. It requires full-duplex connection to be established between sender and receiver prior the data to be transferred. During the connection software of both ends exchange their information in order to verify and ensure an error free communication.

At the both ends, data is sent in stream format. When any program wants to use TCP it will send stream of data to the operating system first then the operating system will send to TCP. TCP will use its internal buffer to store and divide data into packet regardless of the amount of the data. This is the main difference from other protocols in the transport layer, which form packet directly from the data. As a result, TCP will have low overhead and high network throughput. Additionally, TCP uses both MTU of the network and preset duration to define size of the data. The data is received in the same order as it is sent.

During the transmission TCP uses sliding window and positive acknowledgement with retransmission in flow control and error control. The sliding window allows TCP to continuously send the next packet immediately without need to wait for the acknowledge of the previous packet. At the receiver end the same sliding window is used to limit the maximum data that user can send. TCP uses time measurement to check missing or corrupted packets to request a

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 458–462.

retransmission. The positive acknowledgement will send to the sender only when the correct data is received.

TCP of the receiver side inform the size of its buffer to the sender by defining its window size in the acknowledge packet. In a high capacity network the window size will be bigger in order to reduce the idle time that the sender has to wait for an acknowledge. However, a bigger window size will increase datagram and results in increasing network congestion particularly at the connecting point such as router. It will be worse if the router is used between networks that have different speed for example router between Fast Ethernet at the speed 100 Mbps and E-1 at the speed 2.048 Mbps.

3. CONGESTION AND DATA FLOW CONTROL

As the network condition continue to change the TCP have improved its performance in several aspects to suit with both local and remote networks such as round-trip time estimation, slow start and congestion avoidance.

Round-trip time estimation (RTT)

RTT is one of the important parameters of TCP. An underestimate RTT will cause an unnecessary retransmission while an overestimate RTT will delay the packet recovery process. TCP estimate RTT by recording time to receive acknowledgement of previous packets as s1, s2, s3.. Then the Smoothed Round-trip Timed (SRTT) is calculated from

$$SRTT_{i+1} = (\alpha * SRTT_i) + (1- \alpha) * si \qquad (\quad 1 \quad)$$

Where SRTTi is the current RTT. $SRTTi_{+1}$ is the new RTT which will be used for the next transmission. α is a constant between 0 and 1 depending on network status. While the large α will reduce the SRTT dynamic the small α can cause the oscillation. The typical value of α is between 0.8 –0.9.

The retransmission time-out RTO is calculate from

$$RTO_i = \beta * SRTTi \qquad (2)$$

Where β is a constant and greater than 1 in order to adjust RTT. According to TCP (RFC 793) it suggests β=2. Jacobson [6] also suggested that RTT should be collected frequently to achieve a suitable β.

In TCP, one cannot distinguish between an acknowledge of the regular transmission and the acknowledge of the retransmission. This causes the retransmission ambiguity. Phil Kahn [8] proposed an algorithm to improve RTT by the receiver always sending only one acknowledge packet regardless of transmission status and by adding the waiting period of timer back off to the RTT. When the time-out occurs, TCP will increases time-out period to be two times before resending the packet. Thus improving network stability in the situation which the traffic is so congested that missed packet occurs. TCP will adjusts RTO to SRTT once the traffic return to normal situation.

Slow Start

In the packet switching network when the congestion happens each packet will have a longer delay and routers start to drop some packets. As a result the packet lost and more packet being retransmitted and increasing more traffic congestion. The bandwidth reduces to the status called congestion collapse.

To avoid this problem, TCP has to decrease transfer rate once the congestion arises. This calls slow-start and congestion avoidance.

To prevent sending more packets to the congested network, TCP uses a congestion window to set the maximum size of data. It will calculate the minimum between congestion window and advertise window. In a normal situation, the congestion window has the same size as the window of the other end. However, if the waiting period is time-out TCP will interpret as the packet lost due to congestion. It will use slow-start mechanism set congestion window equal to 1 to prevent a retransmission. Once the network resumes normal status i.e. acknowledge packet is sent by the receiver, the slow start mechanism will increase congestion window by 1 for every acknowledge packet it receives. This prevents a sudden increasing in traffic after the network begin to be a normal state or establish a new connection.

Congestion avoidance

TCP on the sender sends only one packet then wait for acknowledge. If the acknowledge is received before time-out, TCP will increase congestion window to 2 and then send 2 packets and wait again. With this technique, waiting time is $\log_2 N$ prior sending N packets.

To prevent a fast increasing in window size that may cause congestion, window size will increase to be 1 as long as every packet in that window is acknowledged. If the condition does not meet, TCP will reduce congestion window to be a half. Therefore window size is decreased exponentially. In general, TCP will reduce transmitting packet to be only 1 packet and wait twice as long as before retransmitting in order to reduce the traffic and gives the router time to process datagram in its buffer.

Slow start and congestion avoidance are not the same mechanism but both are tripped by time-out and able to adjust the congestion window. TCP of the sender set two variables to control the congestion, the congestion window and the threshold.

When time-out occurs, TCP reduces the window size to be a half, keeps it as a threshold and changes congestion window size to be 1 packet. Once the packet is acknowledged, it increases congestion window by 1 if the size is less than the threshold. If the window size is already bigger than threshold, TCP will increase the window size by 1 as long as every packet is acknowledged. Therefore slow-start can quickly set the congestion window. However, when congestion window has its value close to the threshold the congestion avoidance will reduce amount of this increasing.

Currently, TCP assumes that the network is a black block with no any explicit feedbacks to inform the sender about its

congestion. Therefore, the question is how TCP adjusts the congestion window to a suitable size. TCP checks the congestion status by increasing window size until packet lost takes place it then decreases window size. This implicit feedback through the lost of packet results in window size changing back and forth around its stable value. If the window size is too small the network bandwidth will not be fully used, and reduces a system throughput. If the window size is too big, more packets will be lost. Thus it needs a large buffer which in turn increases the waiting time of the job in the system. It may cause some problem in application programs that requires real-time connection or interaction with user like WWW.

TCP mechanisms for control the network congestion

To control the network congestion, TCP adjusts window size according to the congestion status defined as *cwnd(t)*

TCP Tahoe

In TCP Tahoe, the window size *cwnd* is cyclically changed. c*wnd* continues increasing until segment loss occurs. When it does occur, TCP determines that the network is congested, and throttles *cwnd* down to the size of one segment. TCP Tahoe has two phases in increasing *cwnd*; Slow Start phase and Congestion Avoidance Phase. When an ACK segment is received by TCP at the server side at time $t+t_A$, $cwnd(t+t_A)$ is updated from *cwnd(t)* as follows

$$cwnd(t+t_A) = \text{(Slow Start Phase:)} \quad (3)$$
$$cwnd(t)+1, \text{ if } cwnd(t)<ssth;$$
$$\text{(Congestoin Avoidance Phase:)}$$
$$cwnd(t)+1/\, cwnd(t), \text{ if } cwnd(t)>=ssth;$$

where *ssth* is the threshold value at which TCP changes its phase from Slow Start Phase to Congestion Avoidance Phase. When segment loss is detected by timeout or fast retransmission algorithm, *cwnd(t)* and *ssth* are updated as follows;

$$ssth = cwnd(t)/2,\ cwnd(t)=m \quad (4)$$

That is, TCP Tahoe again enters Slow Start Phase when segment loss occurs.

TCP Reno

TCP Reno is similar to TCP Tahoe, but uses another algorithm when segment loss occurs. In Slow-Start Phase and Congestion Avoidance Phase, TCP Reno also use Eq.1 to update the window size, but when segment loss is detected by fast retransmission algorithm, the window size *cwnd(t)* reduces by half. That is

$$ssth = cwnd(t)/2,\ cwnd(t)=ssth \quad (5)$$

Then TCP Reno will enter the fast recovery process. In this step, if the segments of ACK are repeated, window size will increase by 1 segment. The *cwnd(t)* will change to be *ssth* again when there is no repeated ACK after the retransmission. Fig 1 is a typical example of the behavior of *cwnd(t)*

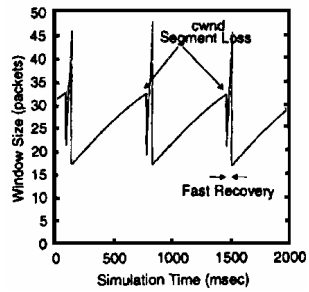


Fig. 1. characteristic of windowsize in TCP Reno

TCP Reno will increase *cwnd* until the packet lost due to a congestion. It will adjust window size which reduce throughput of the network. This process cannot be changed or improved since it is the characteristic of the TCP Reno. Both indicate the network congestion only through the packet lost. This can be a problem because packet lost may be caused by the connection of TCP itself instead of network congestion. However, if the *cwnd* is adjusted properly until there is no packet lost, the throughput will be improved.

TCP Vegas

In TCP Tahoe and Reno, *cwnd* is increased until segment loss occurs due to congestion. Then the window size is throttled, which leads to the throughput degradation of the connection. However, it cannot be avoided because of an essential nature of the congestion control mechanism adopted in TCP Tahoe and Reno. It can detect network congestion information only be segment loss. However, it becomes a problem since the segment may be lost when the TCP connection itself causes the congestion because of its too large window size. If *cwnd* is appropriately controlled such that the segment loss does not occur in the network, the throughput degradation due to throttled window can be avoided. This is the reason that TCP Vegas was introduced. TCP Vegas employs another mechanism for detecting the network congestion. TCP Vegas controls its window size by observing RTTs of packets that the sender host has sent before. If the observed RTTs become large, TCP Vegas recognizes that the network begins to be congested, and throttles the window size. If RTTs become small, the sender host of TCP Vegas determines that the network is relieved from the congestion, and increases the window size again. Hence, the window size in an ideal situation is expected to be converged to an appropriate value and the throughput is not degraded.

In Congestion Avoidance Phase, the window size is updated as:

$$cwnd(t+t_A,) = \begin{cases} \text{cwnd(t)} + 1, & \text{if diff} < (\frac{\alpha}{\text{base\_rtt}}) \\ \text{cwnd(t)}, & \text{if } \frac{\alpha}{\text{base\_rtt}} \le \text{diff} \le \frac{\beta}{\text{base\_rtt}} \\ \text{cwnd(t)} - 1, & \text{if } \frac{\beta}{\text{base\_rtt}} < \text{diff} \end{cases} \quad (6)$$

$$diff = cwnd(t)/base\_rtt - cwnd(t)/rtt$$

where base_rtt [second] is the smallest value of observed RTTs, and α, β are constant.

TCP Vegas has another feature in its congestion control algorithm. That is slow-Slow Start. The rate of increasing cwnd in slow start phase is a half of that in TCP Tahoe and TCP Reno;

$$cwnd(t+t_A) = cwnd(t) + m/2 \quad (7)$$

Note that Eq.(6) used in TCP Vegas indicates that if RTTs of the segment are stable, the window size remains unchanged. Which means the window size is converged to a fixed value in steady state. However, when two or more connections share the bottleneck link, the window sizes are not converged to an identical value. Then, we present and enhance the version of TCP Vegas to prevent the convergence of the window size to a fixed value

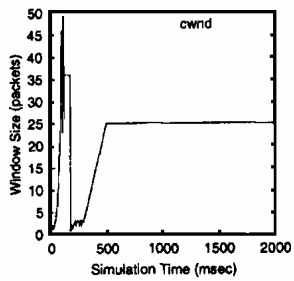


Fig. 2. Characteristic of window size in TCP Vegas

According to this characteristic, TCP Vegas is more efficient than TCP Reno [2]. It has throughput about 31-71% better than TCP Reno and can reduce the retransmission to only 1/5–1/2 times of TCP Reno. Furthermore TCP Vegas employs less data in the network than TCP Reno. As a result it reduces end-to-end delay time. However once TCP Reno and TCP Vegas are used together as a heterogeneous network, TCP Reno shows a better efficiency than TCP Vegas [5].

4. PERFORMANCE AND THROUGHPUT ENHANCEMENT OF TCP VEGAS

In heterogeneous network where both TCP Reno and TCP Vegas are employed in the beginning. TCP Vegas gives a lower throughput than TCP Reno. According to Hasegawa [3] the algorithm used for congestion control in TCP Vegas can be improved either by rapidly increasing window size to be greater than that of TCP Reno or by changing RED algorithm in the router. These will result in TCP Vegas can perform as well as TCP Reno.

In this paper, we propose the technique to improve performance of Hasegawa's TCP Vegas+ [3] based on fast increasing its window size during the beginning phase. This technique will not change the other functions of TCP Vegas, same as Hasegawa's TCP Vegas+. It preserves the advantages that TCP Vegas has in a normal status and improve its performance when it works with other TCP in heterogeneous network. This improved TCP Vegas uses algorithm as follows

- In every incoming ACK segment, the sender will continuously verify window size and RTT. If the current RTT is greater than the previous one while the window size is still the same it means that RTT is increasing due to other TCP connections not the TCP Vegas. Because at that point TCP Vegas increase its window size slower that of other TCP. Therefore, the TCP Vegas must increase it window faster than the other TCP. This can be done by the sender increases its *window size* by twice, to increase the window size rapidly
- If RTT decreases, the sender must reset the *window size* with old value, become to normal state.
- If the packet lost is due to time-out, the *window size* should set to normal. Because the packet lost the TCP Vegas should adjust itself to the normal status. It shows that network congestion may due to the increasing of window size of TCP Vegas itself.
- In normal status TCP improved Vegas will use the same window size a TCP Vegas. However, when the network congested it will use the window size of TCP Reno by checking whether or not the *window size* is at the threshold which is 4.

If the RTT value becomes larger whereas the window size is unchanged, it can be considered that the increasing of RTT is not caused by the TCP Vegas connection itself, but by other TCP connections, which increase its window size more aggressively than the TCP Vegas connection [3]. Then, the TCP Vegas connection should increase its window size rapidly to compete equally or greater than the other connections. When packet loss is detected, the improved TCP Vegas should change its characteristic to the normal state. Because the packet loss indicates the network congestion occurrence, and the congestion may be caused by the rapid increase of the window size of itself.

Network Model

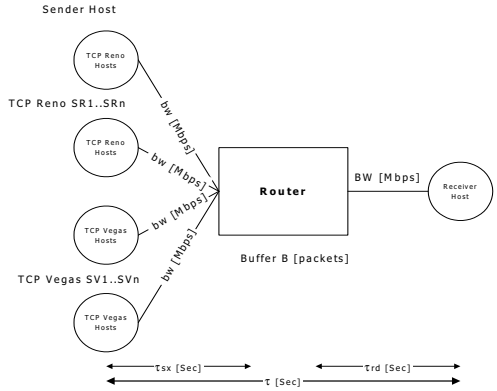


Fig. 3. The used model in the simulations

The used model in this paper is shown in Figure 3. It consists of N_r sender hosts using TCP Reno (SR1,.. SRN_r), N_v Sender hosts using TCP Vegas (SV1,...SVN_v), a receiver host, and router, and links connecting the router and the sender/receiver hosts. The bandwidth of each link between the sender hosts and the router is bw [Mbps]. The bandwidth of the link between the router and the receiver host is BW [Mbps] = μ [packets/sec]. The buffer size at the router is B [packets]. We use the simple model that ignores the scheduling of router buffer. The propagation delay between the sender hosts and the router and that between the router is τsx [sec] and the receiver host are is τrd [sec]. So The total propagation delay between the sender hosts and the receiver host by τ [sec]. The result was show in fig 4

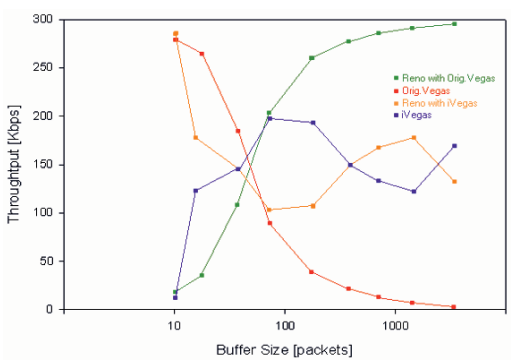


Fig. 4. Simulation result

5. CONCLUSION AND FUTURE DEVELOPEMENT

We have proposed a technique to enhance the performances of TCP Vegas used in heterogeneous network, based on Hasegawa's TCP Vegas+ algorithm. The model and simulation results show that the improved TCP Vegas whose algorithm to control network congestion has been modified yields performances comparable to TCP Reno. However to apply to the real network where the traffic is more complicated and other types of TCP such as TCP-SACK, BIC/CUBIC in Linux [10], TCP–Veno may coexist, more comprehensive models and simulations are needed to be carried out and verified further.

6. REFERENCES

[1] L. S. Brakmo, S. W. O'Malley, and L. L. Peterson, "TCP Vegas: New techniques for congestion detection and avoidance," Proc. ACM SIGCOMM' 94, pp. 24-35, October 1994.

[2] L. S. Brakmo and L. L. Peterson, "TCP Vegas: End to end congestion avoidance on a global Internet," IEEE Journal Selected Areas in Communications, vol. 13, pp. 1465-1480, October 1995.

[3] G. Hasegawa, K.Kurata, M. Murata, "Analysis and improvement of fairness between TCP Reno and Vegas for deployment of TCP Vegas to the Internet," Proc. IEEE ICNP2000, November 2000.

[4] G. Hasegawa, M. Murata, H.Miyahara, "Fairness and stability of congestion control mechanisms of TCP,'' Proc. 11th ITC Specialist Seminar, pp. 255-262, October 1998.

[5] J. Mo, R. J. La, V. Anantharam, and J. Walrand, "Analysis and comparision of TCP Reno and Vegas," Proc. IEEE INFOCOM'99, March 1999.

[6] V. Jacobson, "Congestion avoidance and control," Proc. SIGCOMM'88, Stanford, CA, ACM Computer Communication Review, vol.,18, no.4 , pp.314-329, 1988.

[7] W.Richard Stevens, TCP/IP Illustrated, Volume 1:The Protocols, Addison-Wesley, Reading, Massachusetts, 1994.

[8] Karn, P. and Partridge, C.,"Improving round-trip time estimates in reliable transport protocols," Proc. ACM SIGCOMM'87, 1987.

[9] Tom Kelly, "Scalable TCP: improving performance on highspeed wide area networks," ACM SIGCOMM Computer Communication Review, 2003.

[10] Sangtae Ha, Yusung Kim, Long Le, Injong Rhee, Lisong Xu, "A step toward realistic evaluation of high-speed TCP protocols," International Workshop on Protocols for Fast Long-Distance Networks, February 2006.

Hybrid Scheme by Using Linear Feedback Shift Registers & RSA Security

P.R.Suri, Priti Puri

Department of Computer Science & Applications,Kurukshetra Univesity,kurukshetra,India
pushpa.suri@yahoo.com.puri_priti@rediffmail.com, phone: 09818378953; 011-26472891

Abstract: **This paper starts with stream ciphers based on linear feedback shift registers. A scheme is proposed where both the public key cryptography and the concept of linear feedback shift registers are used. The plaintext message is encrypted with n-Linear Feedback Shift Registers (LFSRs) and secret key, i.e, Initialization vector (IV) of LFSR is being sent with the help of latest public key cryptographic technique, RSA Algorithm. The suggested scheme is successfully applied on text files and on HTML files. The resultant time complexities for different file sizes are shown in tabular format and explained graphically**

I. INTRODUCTION

From e-mail to cellular communication, from secure web access to digital cash, security is an essential art of today's information systems. In security, one important branch is Cryptology. Cryptology is divided into two parts as Cryptography and Cryptanalyst. Cryptography is the art of achieving security by encoding messages to make them non readable. **Plain text** i.e. the original message can be understood by the sender & the recipients**. Cipher text** is codified plain text message using a suitable scheme. The method of producing cipher text from plain text is known as encryption and reverse process is known as decryption.

Cryptography Schemes: There are two types of cryptographic schemes:

1) *Secret key (or symmetric) cryptography (SKC*)

In the symmetric-key approach, both sender and receiver share a common key with a level of trust, is required to ensure that neither party divulges the key. Here, D_k is decryption and E_k is encryption and M is original message.

$$D_k (E_k (M)) = M \quad (1)$$

There are two broad categories of Symmetric algorithms, viz., block cipher and stream cipher. Block cipher transforms a fixed length block of plaintext P_i s into the same length of cipher text C_i s:

$$C = E(P,k) \quad (2)$$

where the encryption function E is applied to P under the influence of a key k of length m bits. Some important block ciphers are used efficiently as DES, 3DES and AES [5].

A stream cipher converts each P_i into C_i as follows:

$$C_i = P_i \text{ XOR } k_i \quad (3)$$

where k_i's are generated through a key stream generator. The security of stream cipher depends entirely on the non-linear structure of the key stream generator. Stream ciphers [2] encrypt individual characters (binary digits) of a plaintext message one at a time using an encryption transformation, which varies with time. Stream ciphers expand a given short random key into a pseudo-random key stream, which is then XOR'ed with the plaintext to generate the cipher text. The used part of the key must be discarded. A seed or initialization vector is used to build the key with a pseudo-random sequence generator.

Stream ciphers are faster than block ciphers with less complex hardware circuitry and also advantageous when transmission errors are highly probable.

There are different types of stream ciphers as synchronous and self synchronising stream cipers and LFSR based stream ciphers.Our concentration is on LFSR based stream ciphers as they can be easily implemented and readily analysed mathematically.

2) Public-key (or Asymmetric) Cryptography (PKC)

PKC use one key for encryption and another for decryption. Whitfield Diffie and Martin Hellman [2] proposed Public Key Encryption in 1976, by this method each user of the network has their own individual private key and public key.

Public Key - k_1, Private key- K_2, Message- M then

$$D_{k2} (E (_{k1} (M)) = D_{k1} (E_{k2} (M)) = M \quad (4)$$

RSA is an algorithm based on public key cryptography; we will use it in our proposed approach.

Hybrid schemes: are those schemes in which Public Key techniques are used for the secret key distribution and once we have securely exchange the key then secret key of symmetric key algorithm is used for encryption and decryption. [5]

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 463–467.

II. PROPOSED APPROACH

The approach utilizes different LFSRs in parallel for encryption and decryption. This approach is using hybrid scheme for security where LFSR is used to encrypt the message & for sending secret key, public key cryptography is used.

The approach first converts the plaintext into binary format and LFSRs start by initialization

Vector values. Here n-LFSR will work together and n-IV functions are required to start n LFSRs.

Initialization Vector of first LFSR, IV_1, starts with one non-linear polynomial function. The variable of this polynomial is behaving like the secret key, which is required at both the sides. For sending the key, the concept of the public key cryptography is used.

The scheme is divided into two parts.

- Working of LFSRs with Intialization Vector generation, create secret key & use this key to encrypt the plaintext messages.
- Sending the secret key at receiver side with the help of public key cryptography.

A. For first part

1) Convert plaintext message into binary format
2) LFSRs start with their IVs.
3) n- LFSRs will work in parellel and for intiate these LFSRs, we required n- IVs.

For Intialization Vector genration

1) Take some non-linear polynomial function (i.e.modulo 2) with a value of variable or indeterminate and output of this polynomial will be IV_1 of $LFSR_1$.
2) For IV_2,value of the variable (X_2) of non-linear polynomial function will be changed and this value will be IV_1 to get the Initialization Vector of the second LFSR and same process will be repeated for $LFSR_3$,$LFSR_4$,....$LFSR_n$.
3) Generated Intialization Vectors, IVs are given to LFSRs and by primitive polynomial concept[1], XOR the different position of LFSRs and get the output of LFSRs.Then this output of LFSRs will be XORed with plaintext binary format and create cipher text .
4) Depending on the position of the characters in the plain text file, XOR the values with the corresponding $LFSR_n$, (where n is between 0-3,fig.2) on the basis of modulo n value. For Example, if the position of some character is 23 then 23 modulo 4 is 3, hence the character is XOR with $LFSR_4$, if it is at position 40 then 40 modulo 4 is 0, hence it is XOR with $LFSR_1$ and so on.
5) After applying steps 3 and 4 on all the characters of the original file, send the new file (Cipher Text) to the other authorized user along with the value of variable of non-linear polynomial used for the determination of IV.

B. For second part

1) Send the value (X_1) of the variable of non-linear polynomial and this will generate IV_1 of $LFSR_1$ at receiving end by using the concept of public key.
2) The value of the variable will be encrypted with RSA algorithm [4] of public key cryptography.
3) Ask for the public key of receiver and encrypt this value with receiver's public key and send it to receiver with encrypted message.

C. RSA Algorithm

The RSA Algorithm [3] [4] is considered to be the most successful Public-Key method ever implemented. Developed by three prominent mathematicians, Ron Rivest, Adi Shamir and Len Adleman (from which the name was derived), the RSA algorithm relied on the integer factorization problem.

1) RSA key generation
Choose two large prime numbers P & Q then calculate

$$N=P*Q \quad (5)$$

Public key (encryption key) E of receiver is not a factor of (P-1) and (Q-1).By this procedure receiver can generate his public key.

Sender will ask for the public key of receiver (encryption key) E to encrypt the value of the variable of non linear polynomial (i.e. secret key) x.

2) RSA Encryption
Cipher text (z) can be created from the plain text x (secret key) as

$$z=x^{E} \bmod N. \quad (6)$$

3) RSA Decryption

At the receiver side, For the private key (decryption key) D of receiver, the following equation is true.

$$(D*E) \bmod (P-1)*(Q-1) =1 \quad (7)$$

Receiver will decrypt the secret key (x) with the help of private key, D as

$$x=z^{D} \bmod N \quad (8)$$

D. *At sending end*

1) The value of variable of non-linear polynomial behaves as plaintext and encrypted with the receiver's public key and sent to the receiver's side.

2) With this encrypted value, cipher text message also sent at receiver side. This cipher text is created by XORing the plain text with the output of the LFSRs. Depending on the position of the characters in the plain text file, XOR the values with the corresponding $LFSR_n$, (where n is between 0-3,fig.2) on the basis of modulo n value.

Initialization Vector generation:

P(X) = Non-linear polynomial

$X=X_1$ (Secret Key)

$Q_1=P(X_1)$

$IV_1=Q_1$ (First IV)

$X_2= IV_1$

$Q_2=P(X_2)$

$IV_2= Q_2$

……..

.....

$X_n= IV_{n-1}$

$Q_n=P(X_n)$

$IV_n= Q_n$

n=Number of LFSRs

E. At Receiving end

1) Receiver will see the encrypted message and encrypted key with his public key.
2) He will decyrpt the value of the variable(X_1) of non -linear polynomail ,by using his private key which is dependent on RSA algorithm.
3) Then apply the value of (X_1) on non-linear polynomial function, to get the Initialization Vector of the first LFSR i.e. IV_1.
4) The same procedure will be repeated on the $LFSR_2$,$LFSR_3$,…$LFSR_n$ as at the encryption side.
5) On the other hand, change the cipher text from the file to its corresponding binary format, character by character. Now depending on the position of the character in the file, XOR the values with the corresponding $LFSR_n$,(where n is between 0-3, fig2.) on the basis of modulo n value.
6) The Plain text file will be obtained after applying steps 4 and 5 on all the characters of the recieved file.

F. Formula representation of the approach

Output of polynomial Q_1 ←Value of variable(X_1) of non linear polynomial $P(X_1)$

$IV_1 \leftarrow Q_1$

$IV_2 \leftarrow Q_2$,output of polynomial $P(X_2)$ with IV_1 as X_2.

$IV_3 \leftarrow Q_3$,output of polynomial $P(X_3)$ with IV_2 as X_3.

…….

…….

$IV_n \leftarrow Q_n$,output of polynomial $P(X_n)$ with IV_{n-1} as X_n.

Binary Text←Plain Text

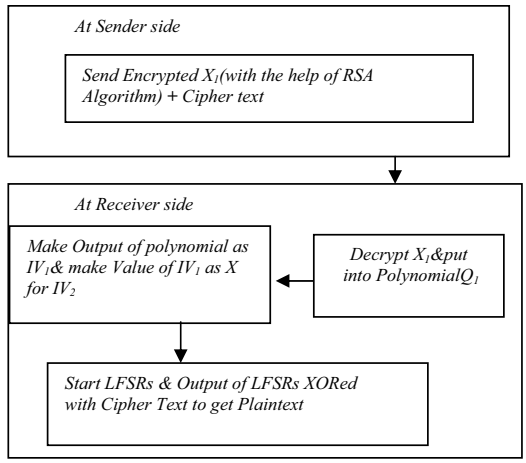


Figure1:IV generation and delivering with RSA Algorithm

Check the value for state modulo n (Fig.2,n=4)

For Remainder = 1 do

PT_1(Binary format) ⊕ $LFSR_1(IV_1)$

For Remainder = 2 do

PT_2(Binary format) ⊕ $LFSR_2(IV_2)$

For Remainder = 3 do

PT_3(Binary format) ⊕ $LFSR_3(IV_3)$

For Remainder = 4 do

PT_4(Binary format) ⊕ $LFSR_4(IV_4)$

……

……For Remainder = n do

PT_n(Binary format) ⊕ $LFSR_n(IV_n)$

Cipher Text ←Binary Text (9)

This gives the Cipher Text to be transmitted.Hence the file is transmitted over the network without any leakage of data.This prevents unauthorized access to the data and ensures security.

III. CONCLUSION

The paper utilizes the concept of hybrid technology where symmetric key as stream cipher is used for encryption and decryption and for sending the secret key, public key cryptography, RSA Algorithm is used. Accordingly the security will be doubled. The complexity of the algorithm will be increased when different polynomials for different LFSRs are used. The values of the IV_2,..IV_n has been depend on the previous values which can be a limitation. This limitation can be overcome by increasing the degree of the polynomials. BRUTE FORCE ATTACK can be avoided here, because first it needs to find the secret key which is encrypted with public key cryptography and then it has to work for Intialization vector polynomial. In this approach, the position of plain text in the file also plays

an important role because XORing of plaintext with output of LFSR depends upon the position of plaintext in the file and that is hidden. The number of LFSR used is also hidden, i.e. the value of n. In the end, we have illustrated a working scheme with four LFSRs and their IV generations diagrammatically. (fig4.) The plaintext byte of index (4i+a), ain (0,…,3), is XORed with the ith byte of the sequence generated by LFSR a. Man in the middle attack can also be avoided because of the use of hybrid key cryptography.

This scheme has been tested for the text and hyper tet files of different sizes with respective encryption and decryption times. The code was found to be working well on Text and HTML files. The source code of the HTML files can be encoded as well, i.e. giving security to the source code of HTML files. The four LFSRs are used; however, these can be increased.

Time Complexities

Implementation times (i.e. encryption and decryption times) of the different files (Text & HTML) are given below with their sizes.

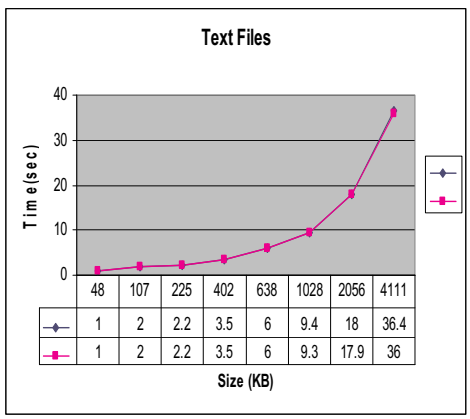


Fig.2.Graphical representation of different text files with their sizes & time complexities

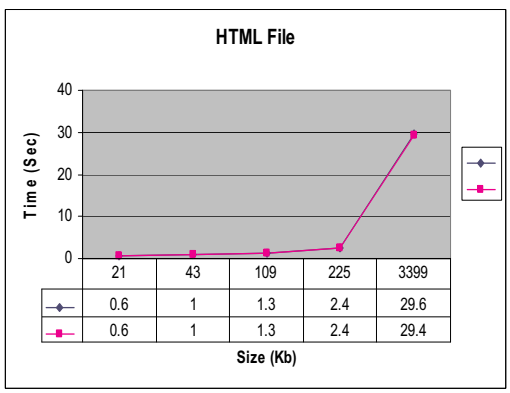


Fig.3.Graphical representation of different HTML files with their sizes & time complexities

Table1 Different Text files with their sizes & time complexities

File Size(KB)	Encryption Time(sec)	Decryption Time(sec)
48	1	1
107	2	2
225	2.2	2.2
402	3.5	3.5
638	6	6
1028	9.4	9.3
2056	18	17.9
4111	36.4	36

Table.2 Different HTML files with their sizes & time complexities

File Size(KB)	Encryption Time(sec)	Decryption Time(sec)
21	0.6	0.6
43	1	1
109	1.3	1.3
225	2.4	2.4
3399	29.6	29.4

REFERENCES

[1] H.Delfs, H.Knebl (2002). “Introduction to Cryptography”, Berlin Heidelberg: Springer-Verlag.

[2] W.Diffie, M. E. Hellman, “New Directions in Cryptography”. IEEE1363 StandardSpecification for Public key Cryptography, 1999. http://grouer.ieee.org/groups/1363.

[3] Matt J. B. Robshaw, “Stream Ciphers Technical Report TR-701”, version 2.0, RSA Laboratories, 1995.

[4] www.rsa.security.com

[5] P. R. Suri,Priti Puri,Application of LFSR with NTRU Algorithm Innovative Algorithms and Techniques in Automation, Industrial Electronics and Telecommunications, Springer Netherlands, 369-373

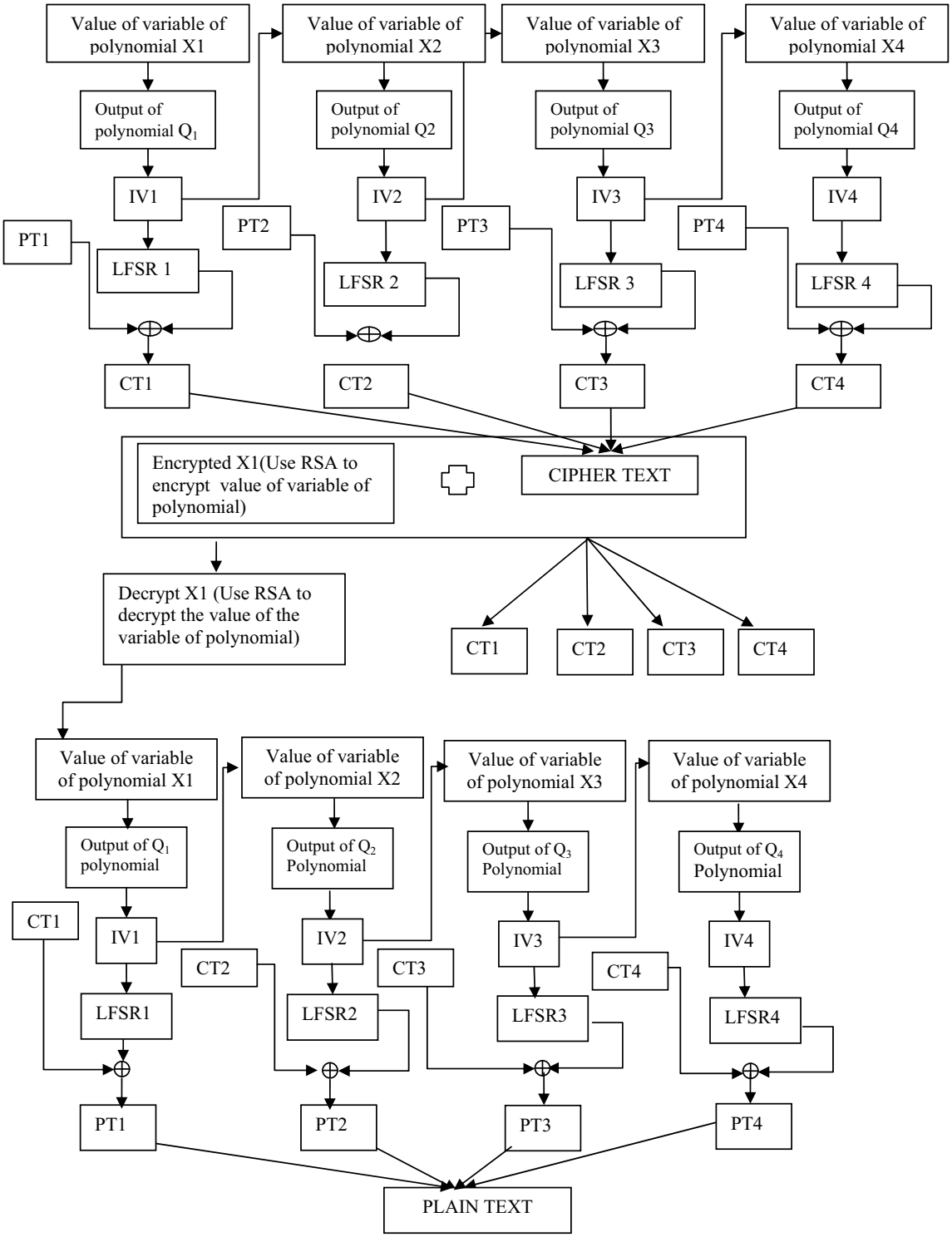


Fig.4: Working scheme with four LFSRs and their IV generation & Security of IVs with RSA.

Analysis of Optical WDM Network Topologies with Application of LRWC Under Symmetric Erlang –C Traffic

Rahul Kundu[1],
[1]*Birla Institute of Technology and Science, Pilani*
rahulkun@gmail.com

V. K. Chaubey[2]
[2] *Electrical & Electronics Group, BITS Pilani*
vkc@bits-pilani.ac.in

Abstract

This paper analyses an Optical WDM network model with application of Limited Range Wavelength Converters for online routing with a wavelength assignment scheme under symmetric Erlang-C traffic. Distinct wavelengths have been assigned to short and long hop connections by wavelength conversion besides attempt to reserve an identical wavelength across all nodes. Probabilistic model for traffic has been developed employing both ring and star topology networks. A mathematical model has been further devised and its call blocking probability has been enunciated. Compared performance analysis of both topologies has been presented with an investigation on collective impacts of various traffic parameters on transmission.

Keywords: LRWC; CR; TDM; WDM; RWA; AON; OPS; Look ahead; Call connection probability

I. INTRODUCTION

The concept of Optical transparent network, based on wavelength division multiplexing (WDM) combined with time division multiplexing (TDM), supported with all optical switching capabilities, has emerged as the most dominating communication technology [1-4]. Optical packet switching combined with WDM technology has changed the static usage of WDM network into an intelligent optical network capable of guaranteeing an efficient routing and switching. [14-16]. In modern optical networks the traffic are provisioned as all optical connection via light paths through intermediate nodes employing efficient routing algorithms[5-8] to enhance the system capacity. The performance of packet switched optical networks are characterized by the strategy involved in handling the packet blocking either through utilizing buffering or wavelength conversion [9-11]. This traffic management requires optical logic processing and data buffering to implement optical burst switching or optical label switching [12-13].

Optical networks using WDM technology have emerged as an attractive solution for meeting rapidly growing demands for bandwidth. Wavelength-routed networks are attractive for realizing the next generation wide-area networks since they offer a data transmission scheme for WDM all-optical networks [18-19]. In wavelength-outed networks, data is transferred on lightpaths. A lightpath is an optical path established between the source node and the destination node. When a connection request arrives, a lightpath is set up. This involves routing and signalling to reserve a wavelength on each link along the path selected [20]. The benefit of wavelength-routed networks is the ability to more fully utilize the bandwidth of optical fibers since they do not require processing, buffering, and opto-electronic-optic conversions at intermediate nodes. It also allows the same fiber to carry many signals independently as long as each uses a different wavelength, known as routing and wavelength assignment (RWA) problem [17]. Calls are additionally subject to the wavelength continuity constraint, which requires a call use the same wavelength on all hops unless wavelength conversion is available at intermediate nodes. If full conversion is available at all nodes, the WDM network is equivalent to a circuit-switched network; however, the high cost of wavelength converters often makes it desirable to keep the amount of conversion used in the network to a minimum. Recent studied have shown that complete or nearly complete set of limited-range wavelength converters (LRWC) [21-22] with limited conversion range (CR) could achieve same packet loss probability as a switch with complete sets of FRWC.

In the present paper we have proposed an appropriate model of WDM optical network using limited range wavelength converters and we have estimated its traffic performance under Erlang-C traffic assuming a symmetric traffic. The network follows a wavelength reservation protocol for wavelength reservation initially and later a scheme for wavelength assignment if the first fails. We have considered both ring and star topology network models to analyse the nature of blocking. The performance response has been measured first against various traffic variables individually. We have further presented a correlative analysis of these parameters and their collective impact on packet transmission rate.

II. PROBLEM DESCRIPTION AND ARCHITECTURE MODEL

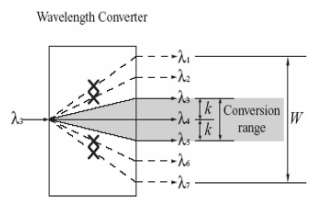


Figure 1: Wavelength Converter

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 468–473.

The present paper is based on general optical network connected in star and ring topologies to serve the optical packet switched traffic. In general, in optical network, the fiber optic links connecting the nodes are unidirectional. The adjacent nodes have two fiber optic cables between them providing the double duplex conditions. Here, every node has employed a packet switch which comprises of a Demux, a Space switch and LRWC, at output [23]. This model intends to follow a wavelength reservation protocol which facilitates reservation of a particular wavelength from the source node itself, thus annulling possibility of wavelength conversion. Though under heavy traffic conditions, it is difficult to always reserve a particular wavelength before packet transmission, this brings in the necessity for wavelength conversion. To cope up with such scenarios when there are no particular wavelength available at all nodes, we have considered a wavelength assignment scheme [20]. This scheme uses a center wavelength for a long hops connection and edge wavelength for short hops connection, each connection request is assigned to wavelength according to its hop number.

Evidently, the noise and delay introduced by multiple channels accumulate to deteriorate the SNR in multiplexed optical networks. Therefore, the objective should be to transfer information over an optical network with minimum number of wavelength conversions. In fact, the network performs satisfactorily below a number of wavelength conversions (*k*) [6].

Further, dispersion, attenuation and cross-talk characteristics of the multiplexed channels ensure that all the channels are not equally efficient. Legal and operational constraints also make the traffic distribution in the channels non-uniform. Even there is a specific band of wavelengths over which the transmission of packets is efficient. Moreover, practical wavelength converters also have few constraints.

Here, we consider a wideband Metropolitan Area Network with a WDM passive star coupler as its physical topology, as shown in Fig. 2, each concentrator is connected through a fiber to the WDM star coupler.

Let ***N*** be the number of concentrators in the system and ***w*** the number of channels wavelength division multiplexed on each fiber. Due to practical considerations in WDM systems, each concentrator is expected to be able to transmit/receive on only a few wavelengths at a time.

For further analysis, we have considered a network connected in ring topology as shown in Fig. 3, there are n nodes named as $N_1, N_2, \ldots N_{n-1}, N_n$. These are connected in series by (n-1) links, which will be henceforth being designated by $l_1, l_2, \ldots, l_{n-1}$. Each link has ***w*** wavelengths multiplexed in each fiber optic link comprising of a single fiber. These wavelengths have been designated as $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \ldots \lambda_w$.

III. MATHEMATICAL MODEL

A. *Star Network*

In order to evaluate performance of a star network, we need to derive the probabilistic evaluation of a star network. The present star topology contains 'n' edge nodes connected via a central hub having λ calls generated at each node with a service rate μ with a channel of capacity C.

Traffic can be expressed as

$$\rho = n\lambda / C\mu \quad (1)$$

The blocking probability for the star topology is derived as for a case of w-channels

$$P_{STAR} = (\rho)^{w+1}\left[1-(\rho)\right] / \left[1-(\rho)^{w+2}\right] \quad (2)$$

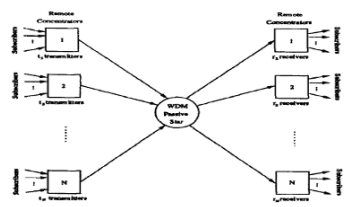


Figure 2: Star Network

The probability that a packet coming at any edge node will be serviced by star node 1-Pstar. This makes the effective number of data packets being routed at each node as λ (1-P_{star}). The Performance parameter for the star topology shown [26] can be written as

$$\rho = 2\lambda(1-P_{STAR}) / \mu \quad (3)$$

and

$$P_{2STAR} = \left[\sum_{k=0}^{w-1}\left(\rho^k / k!\right)P_0\right]^2 \quad (4)$$

Where

$$P_0 = 1 \Big/ \left[1+\rho^w\left(1-\{\rho/w!\}^{w+1}\right) \Big/ (1-\rho/w)w! + \sum_{k=0}^{w-1}\rho^k/k!\right] \quad (5)$$

Net block probability for the star topology can be written as

$$P_{net-star} = (1-P_{STAR})P_{2STAR} \quad (6)$$

Erlang C formula is derived from assumption that a queue is used to hold all request calls which cannot be immediately assigned a channel. The Erlang C formula is given by

$$\text{P}r\text{ [call delayed]} = \frac{A^C}{A^C + C!\left(1-\frac{A}{C}\right)\sum_{k=0}\frac{A^k}{k!}} \quad (7)$$

If no channels are immediately available, the call is delayed and held in a queue and the probability that the delayed call is forced to wait for more than τ_i seconds in the queue is given by, Pr [wait> τ_i | call delayed] = $\exp\frac{-(C-A)}{H}t$ (8)

Where C is the total number of available channels, τ_i is the delay-time and H is the average duration of a call. The probability that any caller is delayed in the queue for a wait time greater than τ_i seconds is given by

Pr [wait> τ_i] = P r [call delayed] Pr [wait> τ_i | call delayed] (9)

$$\Pr[\text{wait} > \tau_i] = \frac{A^C}{A^C + C!\left(1-\frac{A}{C}\right)\sum_{k=0}\frac{A^k}{k!}} * \exp\frac{-(C-A)}{H}t \quad (10)$$

We know from [24] for symmetric traffic assumption, the probability of j packets requesting access to an output fiber in one time slot a_j is given by

$$a_j = \binom{nN}{j}\left(\frac{\rho}{N}\right)^j\left(1-\frac{\rho}{N}\right)^{nN-j} \quad (11)$$

Thus the probability of blocking of packets through a delay-line will be the product of three of the probabilities. It will be given by equation (6), (10) and (11)

$$P_b = P_{net-star} * \Pr[\text{wait} > \tau_i] * a_j \quad (12)$$

$$P_b = (1-P_{STAR})P_{2STAR} * \frac{A^C}{A^C + C!\left(1-\frac{A}{C}\right)\sum_{k=0}\frac{A^k}{k!}} * \exp\frac{-(C-A)}{H}t * \binom{nN}{j}\left(\frac{\rho}{N}\right)^j\left(1-\frac{\rho}{N}\right)^{nN-j} \quad (13)$$

Since there are total p numbers of delay-lines thus the overall probability of blocking will the summation of blocking probabilities due to all the delay-lines, given by

$$P_{block} = \sum_{i=1}^{p} P_b \quad (14)$$

Thus, the probability of transmitting the message must be the complement of the summation of blocking probability possible by all links and channels,

$$P_{trans} = 1 - P_{block} \quad (15)$$

$$P_{trans} = 1 - \sum_{i=1}^{p}(1-P_{STAR})P_{2STAR} * \frac{A^C}{A^C + C!\left(1-\frac{A}{C}\right)\sum_{k=0}\frac{A^k}{k!}} * \exp\frac{-(C-A)}{H}t * \binom{nN}{j}\left(\frac{\rho}{N}\right)^j\left(1-\frac{\rho}{N}\right)^{nN-j} \quad (16)$$

This equation may be used to estimate the total probability of transmission.

B. *Ring Network*

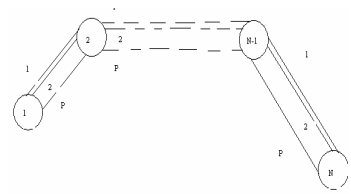


Figure 3: Optical Ring Network

The performance of a network may be evaluated on the basis of its ability to process the arriving packets efficiently. This can be better represented by the parameter ρ, given as $n\lambda/C\mu$.

Now the dropping probability of incoming packets can be written as:

$$P_{RING} = (\rho)^{w+1}\left[1-(\rho)\right] / \left[1-(\rho)^{w+2}\right] \quad (17)$$

The probability that atleast 1 server is free in a link is given as

$$\left[\sum_{k=0}^{w-1}\left(\rho^k/k!\right)P_0\right]$$

So for the entire ring network the probability atleast 1 wavelength is free in all (n/2) links for routing is [26]

$$P_{2RING} = \left[\sum_{k=0}^{w-1}\left(\rho^k/k!\right)P_0\right]^{n/2} \quad (18)$$

The total call connection probability for ring is

$$P_{netRING} = (1-P_{RING})P_{2RING} \quad (19)$$

Hence, similar to (16) transmission probability for a ring topology network could be assigned as:

$$P_{trans} = 1 - \sum_{i=1}^{p}(1-P_{RING})P_{2RING} * \frac{A^C}{A^C + C!\left(1-\frac{A}{C}\right)\sum_{k=0}\frac{A^k}{k!}} * \exp\frac{-(C-A)}{H}t * \binom{nN}{j}\left(\frac{\rho}{N}\right)^j\left(1-\frac{\rho}{N}\right)^{nN-j} \quad (20)$$

IV. Results and Discussions

A. *Star Network*

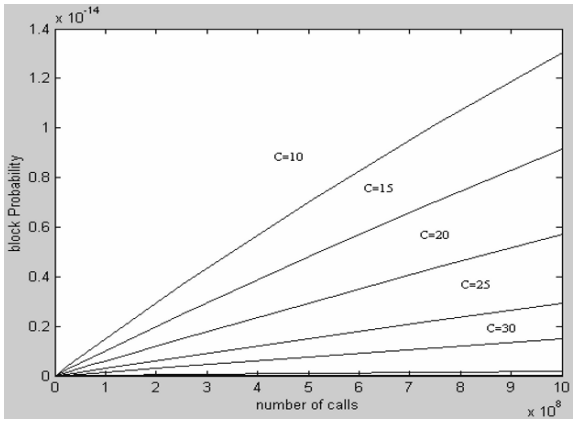


Figure 4: Blocking Probability vs. number of calls

We have analyzed blocking probability characteristics for this particular network model against various parameters of traffic. Network characteristics have been drawn for both topologies. We see that, blocking probability increases proportionally with increasing number of calls, however, blocking decreases for same number of calls when measured across various channels with improving channel capacity, (Fig 4.). This simulation was done for a star network with 5 nodes across channels with varying capacity of 10^10-4*10^10 packet/sec. (Please read the values at multiple of 10^10 in Fig. 4)

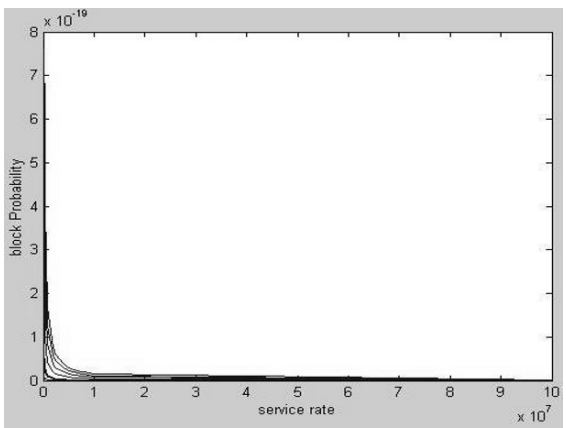


Figure 5: Blocking Probability vs. service rate

Blocking probability shows a response similar to a hyperbolic decay when tested against improving service rate (Fig. 5) however call connection improves with increase in number of calls. We have taken a sample for service rate in range of 10^5-10^7. Number of calls vary within 10^5-10^8. Further investigation shows the network experiences linear increase in blocking with increasing number of nodes (Fig. 6).

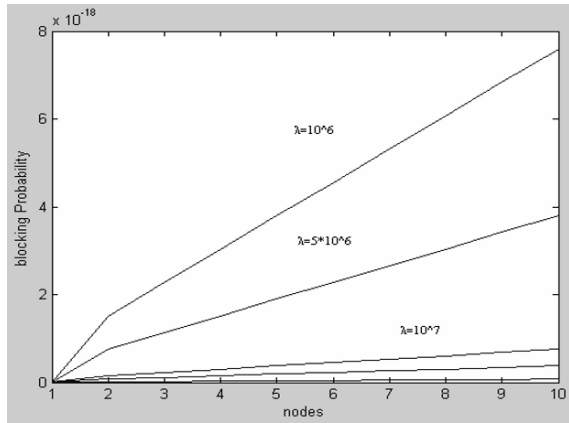


Figure 6: Blocking Probability vs. number of nodes

This network has followed an Erlang-C characteristic which can be best understood with the characteristic [25] shown in Figure 7.

B. *Ring Network*

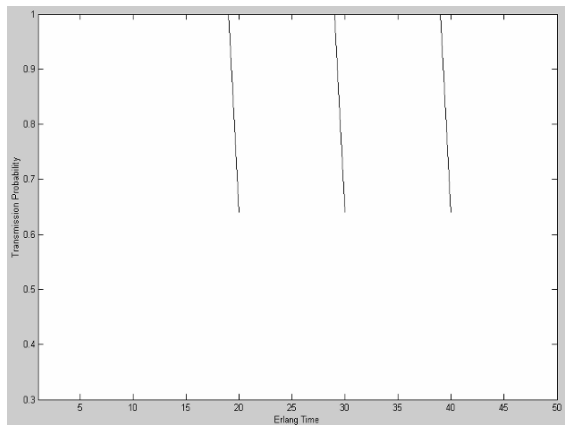


Figure 7: Transmission Probability vs.Erlang C traffic

Probability of blocking decreases sharply with improvement in channel capacity. It goes negligible beyond channels with capacity of 4*10^10. Characteristic has been drawn for increasing values of calls, which has resulted in increase in blocking (Fig. 8) for same channel. The curve between blocking probability and service rate shows a similar response, however blocking probability reduces for the same service rate with improving channels (Fig. 9).

Traffic is largely affected with improvement in channel capacity and increasing service rates. Thus, blocking reduces on improving any of these two parameters. However, blocking

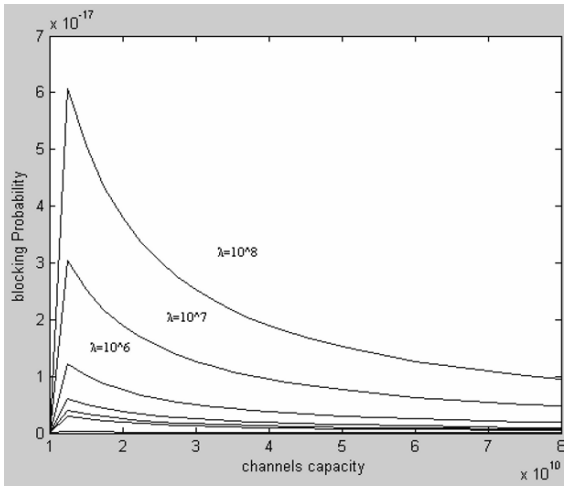


Figure 8: Blocking Probability vs. channel capacity

is affected more by higher service rates as fig 9. Change in blocking probability is more abrupt as shown in the response against service rate than the one with channel capacity (by comparing Fig. 8 and Fig. 9). Thus, in case of a trade off between both parameters, we can yield better traffic performance by optimizing on service rates on cost of channel capacity for Erlang-C symmetric traffic.

Again, blocking shows similar responses against increasing number of calls and higher nodes. We inspect a linear augment in blocking on increasing any of these parameters, depicting almost similar curves for both numbers of calls and number nodes in the network.

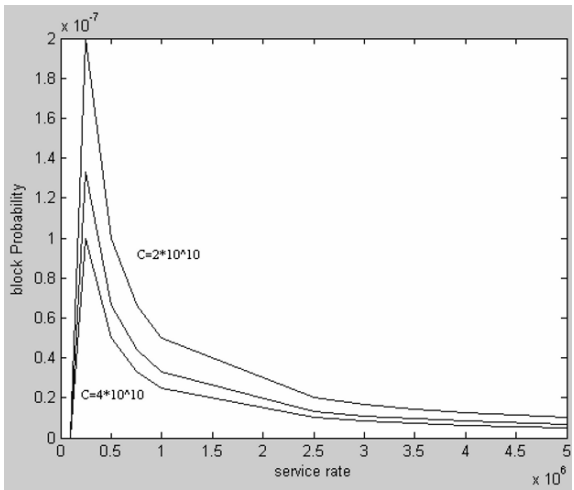


Figure 9: Blocking Probability vs. service rate

V. Conclusion

We have analyzed traffic performance of an Optical WDM network under Erlang C traffic condition. We have deployed Limited Range Wavelength Converters and a wavelength assignment scheme for the purpose of online routing. This assignment scheme is taking care of all wavelength conversions required for packet transmission on the respective nodes, based on their hop connections. However, the network model has a limited application for symmetric traffic. A future work in this regard could be carried out on performance measure of the probabilistic model under other traffic modeling methods like Erlang B which could involve further modifications in the design proposed here. All observations have been conducted for moderate number of nodes. Traffic's Service rate parameter has shown varied response under different topologies. In ring connection, response of blocking for channel capacity and service rate are very similar however is not the case with star connected networks. Which leads us to infer that the service rate behavior has some dependency on network topology also and in our observation; it has exhibited a better performance for star networks. A Further research in this direction could be conducted in analysing other networks like mesh or bus and investigating for the most optimized service rate performance. Other traffic parameters experimented like number of nodes, channel capacity and number of calls, are observed to have exhibited similar responses across both topology networks, we could as well delve into other network topologies to scrutinize whether the nature of performance response curve remains similar across all topologies for these parameters. Blocking has invariably increased with increasing number of nodes and number of calls; otherwise it has gone temperate on increasing channel capacity and service rate, where the latter has exhibited better performance in regard with call connection. The traffic response found from experiments is similar to what is expected in theory. Traffic performance has shown large dependency on number of nodes especially in ring networks, and the performance could also be studied for large ring networks. Experiments have been performed in due consideration of traffic parameter values used in practical purpose and all simulations have been carried out using MATLAB tools and libraries.

References

[1] C.A. Brackett, IEEE J. Sel. Areas Commun. 8 (1990) 948.
[2] I. Cerutti, A. Fumagalli, M. Tacca, A. Lardies, R. Jaganathan, J. Lightwave Technol. 18 (2000) 1649.
[3] G. K. Chang, G. Ellinas, J.K. Gamelin, M.Z. Iqbal, C.A. Brackett, IEEE J. Lightwave Technol. 14 (1996) 1320.
[4] B. Mukharjee, IEEE Network Mag. 6 (3) (1992) 12.
[5] S. Harrish, T.R. Chadramouli, V.K. Chaubey, Indian J. Pune Appl. Phys. 41 (2003) 149.
[6] Abishek Mukherjee, Satinder Pal Singh, V.K. Chaubey, Optical Communications 230 (2004) 59.

[7] G.P. Agarwal, Non Linear Fiber Optics, Academic Press, New York, 1989 (Section 2.3)
[8] G.P. Agarwal, Fiber Optic Communication System, Wiley, New York, 1997, p. 60.
[9] S. Binetti, M. Listani, A. Maga, and R. Sabella "Analysis and dimensioning of switchless Networks for single-layer optical architecture", IEEE JLT 18, (2000) 144-153.
[10] S.J.B. Yoo, "Wavelength Conversion Technologies for WDM Network Applications", JLT, 6, June (1996) 955-966.
[11] J. Diano, and P.L. Chu, "Analysis of partially shared buffering for WDM optical packet switching", IEEE JLT 17, (1999) 2461-2469.
[12] M. Joo and C. Diao, "A new optical burst switching protocol supporting quality of service" Proc, SPIE'98, All optical networks architecture, Control and Management issues, Boston, (1998) 396-400.
[13] John Y Wei, C. D. Lie, S. Y. Park, K. M. Liu, R. S. Ramamurthy, H. Kim, and M. W. Marda, "Network Control and management for the next generation internet" IEICE Trans. Commun. E83-B, (2000) 2191-2209.
[14] F.C.G. Corazza and C. Raffaelli, "An Optical Packet Switch with a Multi-Stage Buffer for IP traffic", Optical Networking, A. Banoni (Ed), Springer, (1999) 300-311.
[15] F. Masetti et al., "High Speed, high capacity ATM optical switches for future telecommunication transport networks", IEEE J SAC, 14, no. 5, (1996) 979-999.
[16] P. Gambini et al., "Transparent Optical Packet Switching: network architecture and demonstrators in the KEOPS project", IEEE J SAC, 16. no. 7, (1998) 1245-1259.
[17] Li-Wei Chen, Eytan Modiano, "Efficient Routing and Wavelength Assignment for Reconfigurable WDM Networks with Wavelength Converters", IEEE INFOCOM 2003.
[18] I. Chlamtac, A. Ganz, and G. Karmi, "Lightpath communications: An approach to high bandwidth optical WAN's," *IEEE Transaction on Communications*, vol. 49, no. 7, pp. 1171–1182, July 1992.
[19] X. Chu, B. Li, and C. Imrich, "Wavelength converter placement under different rwa algorithms in wavelength-routed all-optical networks," *IEEE Transaction on Communications*, vol. 51, no. 4, pp. 607–617, Apr. 2003.
[20] Sho Shimizu, Yutaka Arakawa, Naoaki Yamanaka, "A wavelength assignment scheme for WDM Networks with Limited Range Wavelength Converters", IEEE ICC 2006.
[21] V. Eramo, M. Listani, and M. D. Donato, "Performance evaluation of a bufferless optical packet switch with limited-range wavelength converters", IEEE Photonics Technol. Lett., vol. 16. pp. 644-646 Feb 2004.
[22] V. Eramo, M. Listani, and M. Spaziani, "Resources sharing in optical packet switches with limited-range wavelength converters", J. Lightwave Technol. Vol. 23, pp. 671-687, Feb. 2005.
[23] Reuven E. Gordon, Lawrence R. Chen, "New Control Algorithms in an Optical Packet Switch with Limited-Range Wavelength Converters", IEEE Communications Letters, Vol. 10, NO 6, JUNE 2006.
[24] J. Yates, J. Lacey, D. Everitt, and M. Summerfield, "Limited-range wavelength translation in all-optical networks," IEEE INFOCOMM'96, vol. 3, pp. 954–961, Mar. 1996.
[25] Rahul Kundu, V.K. Chaubey, "Analysis of a CSMA/CA Optical Network with Application of Optical Buffers (FDLs) for Online Routing", IEEE Computer Society, "in press"
[26] Virendra Singh Sheshawat, Dinesh Kumar Tyagi and V K Chaubey, "WDM Network Topologies-A Probabilistic Model", IEEE Computer Society, "in press"

Estimation of Radar Alignment Parameters in Multi Sensor Data Fusion Systems Using MLE Technique

[1]SGK MURTHY,[2]M.V.Ramana Murthy,[3]P.Ram Kumar

[1]
Defence Research & Development Laboratory
Kanchanbagh PO, Hyderabad – 500 058,India
[2]
Department of Mathematics and Computer Science,
Osmania University.Hyderabad – 500 007,India
[3]
Department of Computer Science
and Engineering College Of Engineering,
Osmania University. Hyderabad – 500 007, India

Abstract

Recent trends indicate that multi sensor data fusion is an essential component for C4I applications. Fusion of multiple radar information is needed to ensure more precise target position. To obtain tracking robustness, it is necessary to consider issues that degrade fusion estimation. Systematic errors caused by misalignment of radars degrade fusion estimation. . In multi sensor data fusion, improper alignment of radar sources and geo referencing errors leads erroneous fusion estimate To control such systematic errors , a software is developed by using Maximum Likelihood Estimation (MLE) technique. This paper explains the issues related to systematic errors caused by misalignment of radars and MLE method used for alignment parameter estimation with simulation results.

Keywords: Multi sensor data fusion, Radar tracking, Maximum Likelihood Estimation

Introduction

Multi sensor fusion is a widely used technique in radar tracking applications, in order to obtain the best estimate. Target state vectors obtained from different radars, deployed in different locations are fused to obtain robust and precise estimation of target position[1]. When two radars are observing the same target and its position is measured individually in local co-ordinate frames which are converted into common co-ordinate frame and common time do not match with absolute accuracy. These differences are caused by random errors and systematic deviations due to conversion problems and misalignment of radars in Kalman filter based estimation techniques[3]. Certain physical reasons that create systematic errors are inaccurate geo-referencing of the radars and incorrect setting of horizontal axis and elevation plane of the radars[5] making the fused output erroneous. All these inaccuracies made the fused output

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 474–479.

erroneous. Correction parameters are pre computed by using a common target for correct alignment in real time and better fused estimation of target position. MLE is a statistical technique used for alignment parameter estimation by considering reference and align radar measurement data[6]. Using this technique proto type software is developed and tested with different radar tracks with simulated alignment errors in azimuth and elevation measurements. In the next sections a detailed explanation is given on systematic errors caused by misalignment of radars and MLE technique [2] used to compute the alignment parameters along with software implementation and simulation details. .

2. Systematic errors caused by Radars orientation

Two radars (O_1 and O_2) observe a target located at P_1, and each of them measures it's co-ordinates in own ENV frames. Let the first radar O_1 measure co-ordinates (x_1, z_1) of the target p_1, while the second source O_2 measures the co-ordinates (x_2, z_2) of the same target. Second radar is located in point O_2 and the directions of its coordinate axes do not correspond to the directions of the radar O_1. Due to misalignment the co-ordinates received from the second radar, converted to the coordinate system of the first radar do not match to the true position P_1 [5]. The following diagram (Fig 1) describes the misalignment of radar axis.

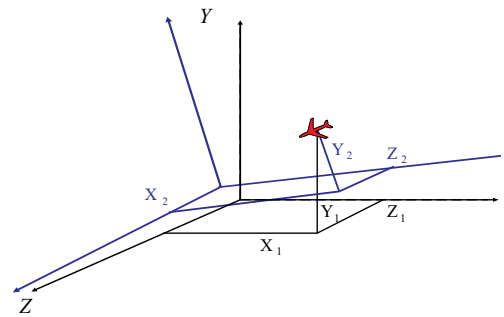


Fig 1

3. Maximum likelihood estimation

The principle of maximum likelihood estimation (MLE) originally developed by R.A.Fisher and it states that the desired probability distribution is the one that makes the observed data "most likely". MLE technique has many interesting properties that make to use in many applications. Mainly it is asymptotically consistent and unbiased; so that for the applications that has larger sample sizes the estimation is converging to true un biased values[3].

For estimating the alignment parameters a model is considered based on the input and out put data. By using MLE technique model parameters are estimated so that the parameter values are most likely to produce the output data.

Let R_A and R_R are align and reference radars located in different locations with their multiple sensors. Tracking the data has to be combined/fused to know the position of the target.

Alignment of the radar R_A is carried out with reference to the radar R_R(reference radar) . Alignment is carried out based on the statistical measurement data of the target, measured from both the radars[7]. Considering geo reference and radar orientation errors a mathematical model is created. The following equation represents the linear statistical model for the problem [8].

$$\begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix} = \mathbf{A} \cdot \begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \end{pmatrix} + \mathbf{b}$$

Where X_1 , Y_1 ,Z_1 be the coordinates of the target measured by the reference radar and X_2 ,Y_2, Z_2 are the coordinates of the target measured by align radar. A and b are the alignment parameters (i.e. rotation and translation matrix) to be estimated. So that the value L where

$$L = \sum_{i=1}^{N} \left(\begin{bmatrix} X_1^{(i)} \\ Y_1^{(i)} \\ Z_1^{(i)} \end{bmatrix} - \mathbf{A} \cdot \begin{bmatrix} X_2^{(i)} \\ Y_2^{(i)} \\ Z_2^{(i)} \end{bmatrix} - \mathbf{b} \right)^T \cdot \left(\begin{bmatrix} X_1^{(i)} \\ Y_1^{(i)} \\ Z_1^{(i)} \end{bmatrix} - \mathbf{A} \cdot \begin{bmatrix} X_2^{(i)} \\ Y_2^{(i)} \\ Z_2^{(i)} \end{bmatrix} - \mathbf{b} \right)$$

Must be minimized and N indicates number of observations measured by both the radars simultaneously. By using log likelihood technique the following matrix equation is obtained which contained 12 unknowns[9] .

$$\sum_{i=1}^{N} \left(\begin{bmatrix} X_1^{(i)} \\ Y_1^{(i)} \\ Z_1^{(i)} \end{bmatrix} - [\mathbf{A}\ \mathbf{b}] \cdot \begin{bmatrix} X_2^{(i)} \\ Y_2^{(i)} \\ Z_2^{(i)} \\ 1 \end{bmatrix} \right) \cdot \begin{bmatrix} X_2^{(i)} \\ Y_2^{(i)} \\ Z_2^{(i)} \\ 1 \end{bmatrix}^T = 0$$

By arranging the terms the solution matrix [A b] is obtained from the above matrix equation. The following equation represents a solution matrix for the problem where [A b] is 3 x 4 matrix. By substituting the input values X_1^i Y_1^i Z_1^i and X_2^i Y_2^i Z_2^i , radar alignment parameters A and b are computed.

$$\begin{pmatrix} \mathbf{A}\ \mathbf{b} \end{pmatrix} = \left(\sum_{i=1}^{N} \begin{bmatrix} X_1^{(i)} \\ Y_1^{(i)} \\ Z_1^{(i)} \end{bmatrix} \cdot \begin{bmatrix} X_2^{(i)} \\ Y_2^{(i)} \\ Z_2^{(i)} \\ 1 \end{bmatrix}^T \right) \cdot \left(\sum_{i=1}^{N} \begin{bmatrix} X_2^{(i)} \\ Y_2^{(i)} \\ Z_2^{(i)} \\ 1 \end{bmatrix} \cdot \begin{bmatrix} X_2^{(i)} \\ Y_2^{(i)} \\ Z_2^{(i)} \\ 1 \end{bmatrix}^T \right)^{-1}$$

4. Software Implementation & Simulation results

4.1 Software Implementation

Using the matrix equation mentioned above, a proto type software is developed under VC++ environment [4] Software accepts Reference and Align radar positions in geodetic system and target measurement coordinates with respect to Reference and Align radars in ENV frame. Alignment parameters with respect to X Y

Z coordinates are computed by pressing the PROCESS Button. XY plane option (Check Box) is used , whenever alignment error is pertaining to azimuth plane only. The following flowchart (fig 2) describes the algorithm used to compute the alignment parameters.

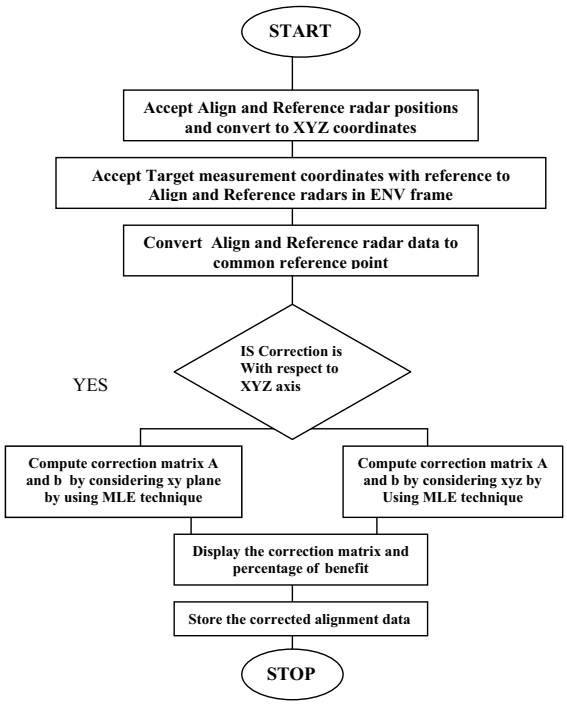


Figure 2

4.2 Input Output Details

Inputs

- Align and Reference radar positions in geodetic coordinate system
- Align and Reference radar measurement data with reference to a common target

Outputs

- Alignment correction matrix A and b
- % of benefit after Alignment correction

4.3 GUI of Radar Alignment software

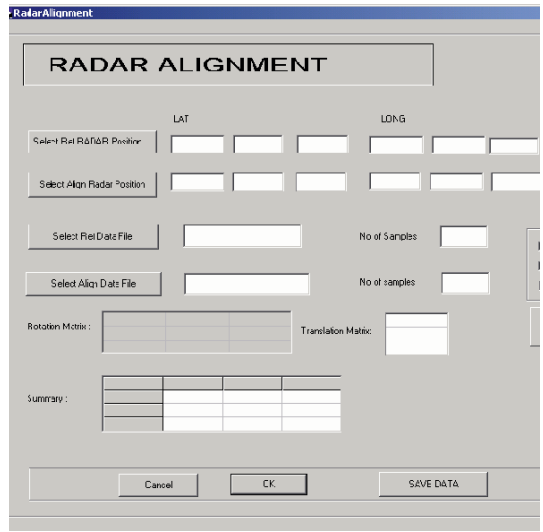


Figure 3

4.3 Simulation Results

Radar alignment Software is tested with simulated data by incorporating systematic errors in azimuth and elevation measurements of the align radar data[10]. It is observed that more noise in the measurement data degrades estimation process. So that some times mean difference after the computation will be more than initial mean

difference. In such a case software indicates 0% improvement and initial coordinate data is retained. The following MATLAB diagrams describe residual values of pre alignment, post alignment and reference data of x y z co-ordinates.

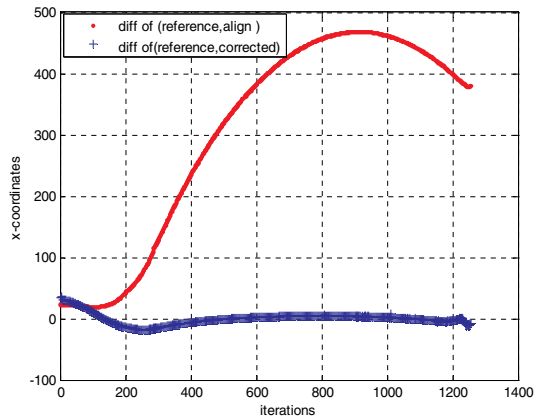


Figure 4

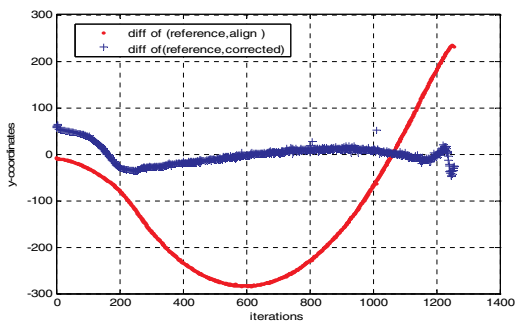


Figure 5

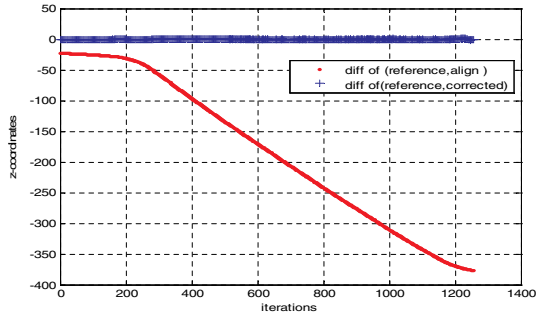


Figure 6

5. Conclusion

In this paper the Novelty of radar alignment process by using maximum likely hood estimation technique is shown. MLE is a statistical parameter estimation technique used to find the model parameters. As it is asymptotically consistent, it is preferred to estimate optimal alignment parameters for a linear statistical model. By using MLE prototype software is developed in VC++ . Software is tested with 5 sets of simulated data by inserting errors in radar azimuth and elevation measurements. However MLE is widely used technique for parameter estimation, it suffers with local minima problem. Certain optimization techniques based on genetic algorithms promise to control local minima problem.

References

1. William Stallings “Data & Computer Communications” Sixth

Edition 2001

2. Myung,jae,(2002), Tutorial on maximum likelihood estimation, Academic press

3. S Blackman,R Popoli,Design and analysis of modern tracking systems,Arteck house,Boston

4. Steven Holzner,Microsoft visual C++ 5,BPB Publications, New Delhi, India

5. Hapgood, M.A,Space Physics Coordinate transformations: A user guide, planet space sci,1992

6. Coordinate systems and transformations, www.spenvis.oma.be

7. David Lee Hall , Mathematical Techniques in multi sensor Data fusion,books.google.com,20048.

8. DL Hall , An Introduction to multi sensor data fusion,Proceedings of IEEE,1997

9. Kosuge , Y Okada, Bias estimation of two 3 – dimensional radars using kalman filter,4th International workshop march, 1996.

10. Bar-Shalom,Y Mobile radar bias estimation using unknown targest Information fusion, Proceedings of 3rd International Conference 2000.

Pre-amp EDFA ASE Noise Minimization for Optical Receiver Transmission Performance Optimization

Akram Abu-aisheh, Saeid Moslehpour
Department of Electrical and Computer Engineering, University of Hartford, 200 Bloomfield Ave, West Hartford, CT 06117 USA
abuaisheh@hartford.edu
moslehpou@hatford.edu

Abstract-Amplified Spontaneous Emission (ASE) noise mitigation from the pre-amp Erbium Doped Fiber Amplifier (EDFA) to the photon detector (PD) in optical receivers can be reduced by optimizing the EDFA at the optical receiver level to achieve optimal optical receiver transmission Performance.

I. INTRODUCTION

The basic design of an optical receiver consists of an EDFA, an optical band pass filter, a photon detector, and an electrical low pass filter. Pre-amp EDFAs are becoming an integral part of optical receivers since their performance is interrelated to the performance of the receiver photon detector. The photon detector used in optical receivers is either a PIN photodiode or an avalanche photodiode (APD). APDs have higher sensitivity than PIN diodes, but they exhibit excess noise that degrades the optical receiver transmission performance. PIN diodes have better noise performance than APDs. So the best optical receiver transmission performance can be obtained by using a combination of a pre-amp EDFA for good sensitivity and a PIN photon detector for low noise.

The generation of noise in doped optical pre-amplifiers is an effect of the spontaneous de-excitation of the laser ions. As the electrons have a finite excited state lifetime, some of the electrons return spontaneously to the ground state emitting a photon. This photon has no coherence characteristics with respect to the incoming light signal, as opposed to a photon generated by stimulated emission. The collection of such spontaneously generated photons, being multiplied by the fiber amplifier, forms a background noise. This background noise is known as amplified spontaneous emission [1]. This is the dominant noise element in pre-amp EDFAs. The optical noise elements of a pre-amp EDFA at different input power, output power, and different signal wavelengths and their effects on the EDFA transmission performance can be analyzed by evaluating the contributions of the EDFA ASE noise generated in the process of signal amplification.

Pre-amp EDFA operation is based on stimulated emission of optically pumped Er +3 ions in silica. Erbium atomic structure and the 3-level atomic level rate equations, for the case of a single stage 980 nm pumped EDFA, were analyzed to help characterize the pre-amp ASE noise. This should help in analyzing the optical receiver transmission performance degradation under different pre-amp operating conditions. The most fundamental limitation to the gain of an erbium-doped fiber amplifier is the energy conservation principle.
This principle can be expressed in terms of the photon flux. The photon flux of the output signal cannot exceed the photon flux of the input signal plus the laser pump photon flux [1].

Since modern receivers use a narrow optical band pass filter at the output of their pre-amp EDFAs, the pump shot noise is filtered in this filter. So pump noise does not play a factor in the pre-amp EDFA transmission performance; however, the pump power affects transmission performance from its contribution to ASE. The most basic treatment of erbium-doped fiber amplifier noise is by analyzing a 3-level erbium atomic system. So the erbium atomic structure needs to be understood to analyze pre-amp noise performance.

II. ERBIUM ATOMIC STRUCTURE

Erbium atomic structure has three energy levels that are of interest for the study of its amplification characteristic for communication use. In 3-three level erbium atomic structure, population inversion can be achieved using laser pumping at 980nm to excite electrons to the upper erbium atomic state. When excited to the upper state, electrons rapidly decay non-radioactively to the meta-stable state. If electrons in the meta-stable state are not stimulated within the electron lifetime in that state, electron transition to the lower states results in spontaneous emission. Spontaneous emission is a random emission that introduces noise. The behavior of this erbium doped fiber atomic structure is described in the following level rate equations [2]:

$$\frac{dN_3}{dt} = -\frac{N_3}{\tau_{32}} + (N_1 - N_3) * \sigma_P * S_P \quad (1)$$

$$\frac{dN_2}{dt} = -\frac{N_2}{\tau_{21}} + \frac{N_3}{\tau_{32}} - (N_2 - N_1) * \sigma_S * S_S \quad (2)$$

$$\frac{dN_1}{dt} = \frac{N_2}{\tau_{21}} - (N_1 - N_3) * \sigma_P * S_P + (N_2 - N_1) * \sigma_S * S_S \quad (3)$$

Where, N is the population density at the given level [1/cm3], S is the photon flux [1/cm2 * s], τ is the spontaneous

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 480–483.

lifetime [s], and σ is the transition cross section [cm^2]. The first equation describes the population change rate for the upper state, the second equation describes the population change rate for the meta-stable state, and the third equation describes the population change rate for the ground state. The steady state atomic populations N1 and N2 are functions of the pumping rate which represents the pump absorption rate between levels 1 and 3, and of the absorption and stimulated emission rates between levels 1 and 2. Figure 1 shows the 3-level erbium atomic structure, and it shows the level transitions when erbium is used in a single stage 980 nm pumped pre-amp EDFA [3]. The sum of the population in the three states of the erbium atomic structure is equal to the total population. The sum of the population in the three states is equal to the total population.

$$N = N1 + N2 + N3 \tag{4}$$

Under a steady state condition, electron transition is given by:

$$\frac{dN_1}{dt} = \frac{dN_2}{dt} = \frac{dN_3}{dt} = 0 \tag{5}$$

The basic principle of signal amplification in erbium doped fiber is based on the fact that when an optical signal passes through erbium doped fiber, the signal is amplified due to stimulated transition between electronic states in the presence of electromagnetic radiation at the correct wavelength to achieve population inversion. In order for signal amplification to happen [4], a frequency f12 is needed where

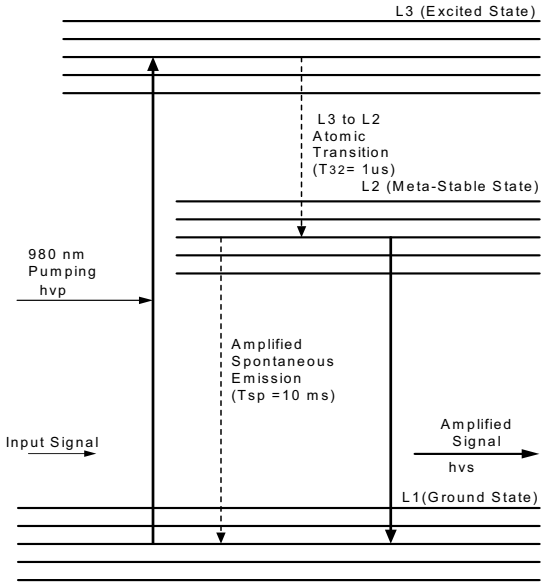


Fig. 1: 980 nm pumping in Erbium atomic structure

$$f12 = \frac{E1 - E2}{\mathrm{h}} \tag{6}$$

Here, h is Plank's constant = 6.626x10E-34 [J/s].
Stimulated photons are in coherence with the input signal, and that results in signal amplification. In free space, the radiation wavelength is given by:

$$\lambda 21 = hc/(E2 - E1) \tag{7}$$

When this radiation interacts with a photon in the lower energy level, the photon is transformed into the upper atomic level. If a photon in the excited state is not stimulated within the 10 ms lifetime of the excited state, it will spontaneously decay to the ground state producing ASE. When this photon travels through the erbium-doped fiber, it gets amplified resulting in amplified spontaneous emission. All the excited ions can spontaneously relax from the upper state to the ground state by emitting a photon that is unrelated to the signal photons. This spontaneously emitted photon can be amplified as it travels down the fiber and stimulates the emission of more photons from excited electrons. Amplified spontaneous emission can occur at any frequency within the fluorescence spectrum of the amplifier transitions. The dominant noise source in any EDFA is amplified spontaneous emission. This spontaneous emission reduces the amplifier gain by consuming the photons that would otherwise be used for stimulated emission of the input signal. In single mode fiber, the noise output power resulting from amplified spontaneous emission is given by [5]:

$$P_{ASE} = 2 * n_{sp}(G - 1)hv\,\Delta(v) \tag{8}$$

The total amplified spontaneous emission at any point in the fiber is the sum of all amplified spontaneous emission power from the previous sections in the fiber and the amplified spontaneous emission at the given fiber point.

In order to minimize ASE noise, the pump power should be just enough to achieve population inversion. Population inversion can be achieved when the population in the excited state, N2, is greater than the population in the ground state N1. The threshold pump power required to achieve population inversion can be obtained by setting the rate equation of level 2 to 0, and setting N1 to be equal to N2.

A long meta-stable state lifetime and a large absorption cross section are needed to have a low pump threshold to achieve population inversion. A detailed analysis of EDFA and photodiodes noise elements was performed by different researchers [6] & [7]. When the EDFA ASE noise is converted to an electrical signal in the photodiode, it gives rise to signal spontaneous beat noise and spontaneous-spontaneous beat noise. When pre-amps are used in the optical receiver, signal spontaneous beat noise dominates all other noise elements in the photon detector [8].

III. Optical Receiver Transmission Performance

Optical receiver transmission performance, commonly known as bit error rate (BER) performance, is the gauge by which optical receivers are characterized. It characterizes the ability of the receiver to perform up to the transmission performance specifications under the same test conditions as those where the receiver operates in the field [3]. So transmission performance will be used to analyze the optical receiver performance under different operating conditions. The setup of figure 2 was used to perform the optical receiver tests analyzed in this paper. This setup has the ability and flexibility to perform optical receiver transmission performance measurement, optical input and output measurements, optical signal and noise control, and optical input SNR control. Since the most important factor in the pre-amp performance is how well it performs in the optical receiver, the optical receiver optimal transmission performance analysis under different operating conditions is the ultimate method for characterizing pre-amp EDFA noise performance. The pre-amp EDFA design needs to be optimized at the pre-amp level and the EDFA level. Then the pre-amp performance should be determined by how well the pre-amp performs in the optical receiver.

For optimal optical receiver transmission performance, the pre-amp EDFA design must be coordinated with the photon detector design in order to minimize amplified spontaneous emission noise mitigation from the pre-amp EDFA to the photon detector and the photon detector signal-spontaneous beat noise. The pre-amp input power, output power, and operating wavelength should be taken into account. This allows designers to choose the right erbium doped fiber length and pump power combination, and it helps minimize amplified spontaneous emission at the output of the pre-amp EDFA.

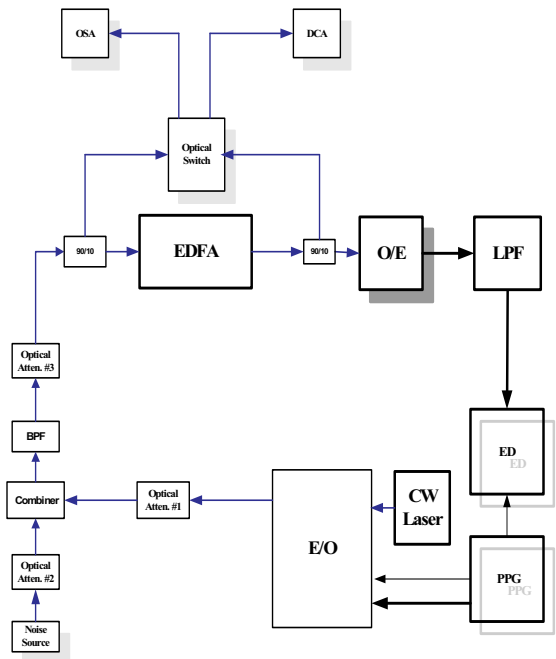


Fig. 2: Optical receiver transmission performance test set-up

Since the most important factor in pre-amp performance is how well it performs in the optical receiver, optical receiver optimal transmission performance analysis under different operating conditions is the ultimate method for optimizing pre-amp EDFA performance in the optical receiver. The pre-amp EDFA design needs to be optimized at two levels: the pre-amp/photon detector subsystem level and the optical receiver level.

Several characterization experiments were performed to analyze the effects of changing the pre-amp operating conditions on the optical receiver transmission performance. Testing the pre-amp-based optical receiver at a fixed signal-to-noise ratio at 1550 nm, the transmission performance was recorded at different input/output combinations. The optical system transmission performance results are given in Table I.

TABLE I
OPTICAL RECEIVER PERFORMANCE AT DIFFERENT INPUT AND OUTPUT POWERS

System BER	Pin=-24dbm	Pin=-22dbm	Pin=-20dbm
Pout=-11.5dbm	4.30E-08	8.90E-09	3.10E-09
Pout=-10.5dbm	4.00E-09	7.40E-10	3.20E-10
Pout=-9.5dbm	3.60E-10	1.40E-10	6.20E-11
Pout=-8.5dbm	1.80E-10	7.60E-11	4.80E-11
Pout=-7.5dbm	1.40E-10	6.50E-11	4.60E-11

A graphical representation of the optical system transmission performance results, after normalizing BER, is given in Fig. 3.

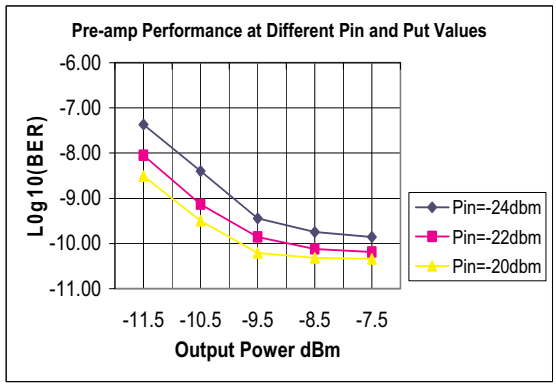


Fig. 3: Optical receiver performance change at different input and output power levels

From the results in figure 3, we see that the optical receiver transmission performance improves as the pre-amp output power is increased. This improvement is due to the fact that more output power requires more pump output, and more output power excites more electrons to the upper state. This excitation will result in the population inversion that is needed for the amplification process.

Testing the pre-amp-based optical receiver at different input powers and at different signal to noise ratios at fixed output power and input signal wavelength, the transmission performance changes due to the changes in the operating conditions were monitored, and the results are given in Table II. The results given in Table II show that the optical receiver transmission performance improves as the pre-amp input signal power is increased.

TABLE II
BER AT DIFFERENT INPUT POWERS AND SIGNAL TO NOISE RATIOS

System BER	SNR=8dB	SNR=9dB	SNR=10dB	SNR=11dB	SNR=12dB
Pin=-28dBm	2.5E-06	3.9E-07	7.6E-08	1.2E-08	2.3E-09
Pin=-27dBm	7.7E-07	8.2E-08	1.3E-08	9.5E-10	1.2E-10
Pin=-26dBm	2.3E-07	2.0E-08	1.6E-09	9.0E-11	7.6E-12
Pin=-25dBm	8.5E-08	4.9E-09	3.1E-10	1.3E-11	6.0E-13

A graphical representation of the system transmission performance, after normalizing BER, is given in Figure 4.
At the atomic structure level, an increase in the input power causes more stimulated emission of the excited electrons. This leaves fewer electrons to move to the ground state spontaneously. This means that the pre-amp is generating less amplified spontaneous emission, and this reduces the signal spontaneous noise in the photon detector which results in improved optical receiver transmission performance.

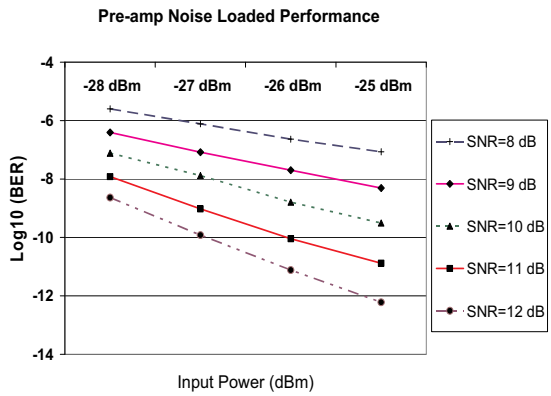


Fig. 4: Optical receiver performance change at different input powers and different input signal to noise ratios

IV. CONCLUSION

Since the most important performance factor in the pre-amp performance is how well it performs in the optical receiver, the optical receiver optimal transmission performance analysis under different operating conditions is the ultimate method for characterizing pre-amp EDFAs noise performance. The results of the tests performed for this work shows a need for designing the pre-amp EDFA and photon detector as one subsystem. Then the pre-amp EDFA needs to be fine-tuned at the optical receiver level to achieve optimal optical receiver transmission performance.

For optimal optical receiver transmission performance, the pre-amp EDFA design must be coordinated with the photon detector design in order to minimize amplified spontaneous emission noise mitigation from the pre-amp EDFA to the photon detector. This will minimize the photon detector signal-spontaneous beat noise. Both the pre-amp EDFA and the photon detector should be treated as one block during the design stage.

REFERENCES

[1] E. Desurvire. "*Erbium-Doped Fiber Amplifiers, Principles and Applications"*, Wiley, NY, (1994)
[2] J. T. Verdeyen, 1995, *"Laser Electronics"*, third edition, Prentice Hall.
[3] A. Abu-aisheh, 2003, "Pre-amp EDFA Noise Characterization for Optimal Optical Receiver Transmission performance," *Ph.D. dissertation*, Dept. of Elec. Eng. Florida Institute of Technology, Melbourne, FL.
[4] J. Gowar, 1993 *"Optical Communication Systems"*, second edition, Prentice Hall Inc.
[5] O. Becker and Simpson, 1999 *"Erbium-Doped Fiber Amplifiers, Fundamentals and Technology"*, Academic Press, NY.
[6] R. Tucker and H. Kingston " Optical Sources Detectors and Systems Fundamentals and Applications" Academic Press, Inc, (1995)
[7] R. S. Tucker and D. M. Baney, 2001 "Optical Noise Figure: Theory and Measurement" OFC, Anaheim, CA.
[8] S. Soerensen, (2000*) Journal of Light Wave Technology*, "Optical Beat Noise Suppression and Power Equalization in Sub-carrier Multiple Access Passive Optical Networks by Downstream Feedback" VOL. 18, NO 10.

Light Weight Cryptography and Applications

Sandeep Sadanandan[1], Rajyalakshmi Mahalingam[2]

[1]*Department of Computer Science, Technical University Munich, Munich, Germany.*

sandeep.sadanandan@in.tum.de

[2]*Technical University Munich, Munich, Germany*

rajee.mahal@mytum.de

***Abstract*— Portable handheld devices are emerging as strong competitors to large traditional electronic devices requiring advanced safety and security mechanisms. Lightweight cryptography is developing as an attractive public-key cryptosystem for use with many such devices, particularly when the devices are not capable of performing sufficient cryptographic operations to guarantee adequate security. The paper provides a survey, how Elliptic Curve Cryptography as a light weight security solution, is accommodating the limitations of resource constrained environments of a whole lot of new generation applications.**

I. INTRODUCTION

The field of cryptography, which started as a simple encryption technique has broadened to areas like authentication, data integrity, confidentiality and non-repudiation of information. It is no more a surprise to anyone that new forms of cryptography came soon after the widespread development of electronic communications.

Some cryptographic methods rely on the secrecy of encryption algorithms. Such methods are now only of historical interest and are not sufficient for real-world needs. All modern algorithms base their security on the usage of keys instead of the secrecy of the method. Symmetric key cryptography uses a single key for both encryption and decryption, thereby inheriting the name called private key cryptography. With this form of cryptography, it is obvious that the key must be known to both the sender and the receiver, which in fact, is the secret. The biggest difficulty with this approach, of course, is the distribution and management of the keys. If there are *n* persons who want to communicate with one another, each of them have to know the private key of every other person. This means that all the *n* persons would have to safely keep *n* keys each.

Asymmetric key cryptography also known as Public Key Cryptography, which evolved to address the key management issues posed by symmetric cryptography is said to be the most significant development in cryptography. This cryptosystem employs two keys instead of a single secret key. One of the keys is freely distributable, called public key and is used for encryption. The second key is the secret or private key and is used for decryption. As the private key cannot be computed from the public key, the system remains secure even though one of the keys is made public. Each person has to keep only his own private key, which solves the problem of key management. Additionally, unlike the symmetric key cryptosystem, public key cryptography facilitates a method for generating digital signatures, there by addressing the problem of guaranteeing unique authentication and non-repudiation.

Strong cryptographic algorithms are designed to be executed on specialized hardware devices or computers. However, the latter point is becoming less significant due to the increase in the computing power of computers. As a result, the algorithms started acquiring longer keys to provide adequate protection. Generally, symmetric algorithms are much faster to execute on a computer than asymmetric ones. In practice, they are often used together, so that an asymmetric key algorithm is used to exchange a randomly generated encryption key, and the generated key is used to encrypt the actual message using a symmetric algorithm.

Today, the wide variety of devices in the digital world ranges from desktops to mobile phones. Equipments such as laptops, handhelds, embedded processors and smartcards have been an incalculable boon to the expanse of electronic society. These devices are prized for the flexibility and conveniences they provide while simultaneously being very demanding in terms of security to avoid intrusion and tampering. The main challenge, therefore, is to develop new efficient cryptographic techniques capable of optimally fitting in various platforms and provide an optimized cost/performance trade-off.

Existing methods of security are not sufficient as they are generally software dependent to carry out thousand bit operations. Software implementations need more resources and time, hence are quite expensive. What we actually need is to implement the computationally intensive operations with limited memory and power at an acceptable speed without compromising on security. This is where the new field of *Light Weight Cryptography* (LWC) emerges. *Elliptic Curve Cryptography* (ECC) [1] based on the algebraic structure of elliptic curves over finite fields[1] is one such approach for light weight cryptographic solutions. It is the most rich of all the light weight cryptosystems at present and could be considered as a synonym for LWC. Section II presents the mathematical preliminaries of the cryptographic algorithms.

The smaller key size in ECC makes it possible to have more compact implementations for a given level of security, which means faster cryptographic operations, running on smaller chips or more compact software. Section III examines as to why ECC system is most suitable over classical methods. An

[1] A finite field is a closed finite set under two binary operations; satisfying associativity, having identity elements and all elements having inverses wrt. both the operations.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 484–488.

overview of current ECC applications is given in Section IV followed by the conclusion in the end.

II. Cryptographic Arithmetic

Traditional public key cryptosystems base their security on hard to solve mathematical problems of Integer Factorization Problem (IFP) or Discrete Logarithm Problem (DLP) in finite fields.

The general integer factorization problem involves the factorization of a given positive integer n for two or more prime factors, i.e., finding p and q such that $n = p \times q$. Rivest Shamir Adleman (RSA) is the best known cryptographic algorithm that bases its security on the difficulty of solving IFP. Coming to discrete logarithm problem, the method deals with a group $G = Z_p^*$ with p-1 elements. Given the prime p, a generator[2] α and an element β of G, DLP aims at finding the unique integer x, $0 \le x \le p - 2$ such that $\alpha^x = \beta \pmod p$. Many cryptographic applications including Diffie-Hellman (DH) key agreement, El-Gamal encryption and digital signature algorithm (DSA) are based on DLP. These problems are unbelievably hard to solve. For an input of a 1024-bit integer, it could take millions of years to find the solution.

In 1987, Neal Koblitz [1] and Victor Miller independently came up with the concept of Elliptic curve cryptography (ECC). They are DLP based systems and therefore their security is based on the difficulty of the *Elliptic curve Discrete Logarithm Problem* (ECDLP). But unlike DLP, there are no sub-exponential time algorithms to solve ECDLP and therefore is stronger. The elliptic curve cryptosystem operates in a group of points of an elliptic curve defined over a finite field. Most ECC implementations use special types of finite fields for good performance, namely, binary extension fields F_2^m, prime fields F_p or generally F_p^m.

The application of elliptic curves for Diffie Hellman key agreement scheme is detailed in Fig. 1. The security of the system lies in the fact that, even if an eavesdropper manages to get Q_a (or Q_b), she cannot retrieve the secret number A of Alice (or B of Bob). Calculating this value is practically infeasible which is guaranteed by ECDLP.

As can be seen from Fig. 1, the basic operation of all ECC systems is scalar multiplication. The operation of computing $kP = P + P + ... + P$ is called scalar multiplication of a point P, a two-dimensional point on the elliptic curve E. More specifically, public keys are created from these multiples of the generating point, therefore, in our case, $Q = kP$ is the public key for the private key k with (k, Q) as our key pair. ECC exploits the fact that given the points P and $Q = kP$ on the curve E, it is practically impossible to calculate k.

Fig. 2 depicts the addition of two points geometrically. If M and N are two points of an elliptic curve, a line is drawn passing through the points M and N. A special property of elliptic curves guarantees that the line will also pass through a third point, which is shown as $-T$ in the figure. The sum of the points M and N is defined to be the reflection of the point $-T$ on the x-axis, which is the point T in the figure.

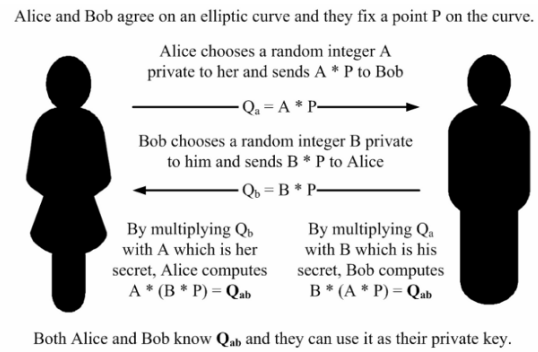


Fig. 1. Diffie Hellman Key Agreement using Elliptic Curves.

Researchers have come a long way from Koblitz curves [1] to Edwards coordinates [2] in speeding up the operation of point addition.

Soon after the introduction of ECC, in 1989 Neal Koblitz [3] proposed the idea of cryptography with generic *Hyperelliptic Curves* (HECC) over a finite field. Hyperelliptic curves are a special class of algebraic curves and can be viewed as generalizations of elliptic curves. A hyperelliptic curve of genus g = 1 is an elliptic curve. Elliptic curves have been extensively studied for over a hundred years, whereas the theory of hyperelliptic curves has not received as much attention by the research community.

In HECC, the addition methods vary from the complex generic algorithm by Cantor [4] to Gaudry's [5] addition formulae for a special type of hyperelliptic curves. Instead of the addition of points as in ECC, pairs of points are combined together in HECC. Anyhow, the overhead of more complicated algorithms in HECC is compensated with the added security provided by HECC.

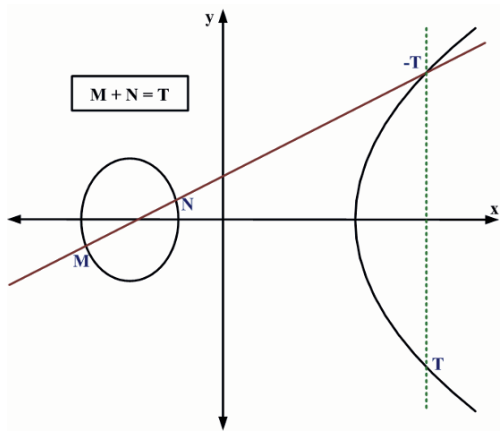


Fig. 2. Addition of points on an elliptic curve.

[2] A generator is an element of the group which can generate the whole of the group by repeated addition/multiplication.

III. ECC Gaining

While comparing different cryptosystems, there are a few primary things of focus like security, speed and cost. These comparisons, pave the way for the cryptographic standards to be used in the coming decades.

A. Security

The first and foremost advantage of ECC/HECC compared to RSA is the absence of sub-exponential algorithms to solve ECLDP. As a consequence, ECC uses much shorter keys to achieve the same security level, resulting in faster and smaller signature and certificate implementations.

Over the past decade, innumerous comparisons were made between RSA/DH and ECC on the level of security and key size. Table I presents the results and the key sizes recommended by National Institute of Standards and Technology (NIST) for different cryptographic algorithms [6].

As the level of security requirement goes high, RSA/DH has their key-sizes growing much faster than ECC. Clearly, ECC provides more security per bit increase than the available traditional cryptosystems. More precisely, when talking about security, ECC is definitely far ahead than the other asymmetric algorithms.

B. Speed

The evident difference in the key sizes directly contributes to the differences in efficiency and speed of respective implementations. The elliptic curve computations are more complex than RSA/DH, but the security achieve per extra bit makes up for the extra arithmetic.

The mobile and wireless devices of the present age have small word-size for processor, normally 16 or 32 bits. The cryptographic algorithms implemented on them usually require very long operands – ranging from 163-521 bits for ECC and 1024-15360 bits for RSA/DH.

TABLE I
NIST Recommended Key Sizes

Bits of Security	Symmetric Key Algorithm	RSA/DH (key size)	ECC Field (size)	
			Prime F_p	Binary F_2^m
80	2TDEA	1024	192	163
112	3TDEA	2048	224	233
128	AES-128	3072	256	283
192	AES-192	7680	384	409
256	AES-256	15360	521	571

It has been proven [8] that the relative performance of ECC systems over RSA systems increases as the word size of the processor decreases. This stems from the fact that the complexity of addition/subtraction and optimized reduction based on sparse Pseudo-Mersenne-Primes in ECC grows linearly with the decrease of the word size. At the same time Montgomery reduction of RSA grows quadratically with the decrease of the word size.

The comparative studies done in [9] shows that Elliptic Curve Discrete Signature Algorithm (ECDSA) [6] offers better performance over RSA and in significant overall timing. The Signing/Verifying Performance of ECDSA versus RSA measured on Windows XP with an Intel 3.00 GHz Pentium 4 processor, and 512MB of memory, using Security Builder by Certicom toolkits [7] is given in Table II.

The bandwidth required to perform key agreements and digital signatures on a communication link is closely related to the key size of different public key systems. The key sizes of cryptographic algorithms given in Table I also reflect roughly the number of bits that need to be transmitted each way over a communication channel for a key exchange. Not just limited to devices with low computing power, ECC offers a much better solution than first generation public key systems in channel-constrained environments too.

C. Real-Estate and Power

ECC offers a multitude of implementation options, both on software and hardware. The number of gates [3] used for implementing a cryptographic algorithm in hardware signifies the gate-equivalent of the algorithm. Gate equivalent is also considered as an indicator of the size and power consumption of the respective hardware implementation. In small scale devices, where power and real-estate are of utmost importance, even a small reduction in gate-equivalent of a circuit accounts to a considerable achievement.

Comparing with symmetric cryptosystems, ECC still requires about 5-10 times more gates for hardware implementation. A minimum gate-equivalent of 10113 gates has been achieved for 113 bit ECC crypto processor in [10]. At the same time, the block cipher 'PRESENT' [11] is implemented with a gate-equivalent of 1200 gates for a maximum key length of 128 bits. The authors of [12] have estimated that instances of ECC and HECC for secure identification protocols could be implemented requiring less than 5000 gates.

TABLE II
Signing/Verifying Performance of ECDSA versus RSA (Certicom) [7]

Key Size (bits)		Symmetric Key	Sign (Sigs/Min)		ECC Benefit (%)	Verify (Sigs/Min)		ECC Benefit (%)
ECC	RSA		ECDSA	RSA		ECDSA	RSA	
224	2048	3-DES	105840	2940	3600	47520	26880	177
256	3072	AES-128	54000	480	11250	22800	11280	202
384	7680	AES192	30960	60	51600	11040	2160	511
521	15360	AES-256	14400	60	24000	5280	480	1100

TABLE III
RELATIVE COMPUTATION COST OF DIFFIE-HELLMAN AND ELLIPTIC CURVES

Security Level (bits)	Ratio of DH Cost : EC Cost
80	3:1
112	6:1
128	10:1
192	32:1
256	64:1

The systematic evaluation of ECC crypto-processor on RFID [13] also shows that ECC could meet the real estate size requirements of RFIDs at the cost of some reduced functionality. However, the authors in [13] argue that the result is still reasonable for RFIDs with reduced power consumption and at a clock frequency of 175 *kHz* on 180 *nm* technology.

A comparative study of the cost incurred by DH and ECC computations has shown that ECC is much cheaper than DH. Table III [14] shows results of the comparison for the NIST recommended key-sizes in Table I.

D. Hyperelliptic Curves – Closing the Gap

We have already seen that the reduction in the number of bits leads to lesser computational operations and is adequately suitable for low bandwidth scenarios. When we go to higher genus hyperelliptic curves, it is shown that we need less number of bits for the same level of security. Table IV gives a comparison of the bits needed for curves of different genus [15].

As the genus of the curve goes higher, the arithmetic of encryption gets complicated. Even in the simplest genus 2 case, the basic operation of scalar multiplication needs 25 computer multiplications and 1 division, whereas it takes less than 16 multiplications in the case of ECC [15]. Although in recent years [5], better arithmetic has been shown for hyperelliptic curves. Researchers of the (hyper)elliptic curve cryptography field have still not reached a consensus on which system is better [16][17].

As the higher genus curves with genus 3 and 4 suffer from some security attacks, curves of genus 2 are the focus of research community. Basically, there has to be a catch behind the fact that hyperelliptic curves of higher genus are still in the background. The catch is that deciding whether a hyperelliptic curve is secure or not, is still very slow. Also, the group additions have to be speeded up to match ECC.

IV. APPLICATIONS

Over the past few years, the rapid growth in the use of networking devices has led to many pervasive smaller and useful devices. Without digging too much we can find numerable examples around us such as PDAs, Mobile phones, computer gadgets, automotive remote controls and other consumer services like online payments, sensor applications, smart cash etc.

[3] Gate: The basic building block of any electronic circuit (NAND/NOR etc.).

TABLE IV
ECC VS. HECC

Genus	Bits Needed	Number of	
		Multiplications	Divisions
1 (ECC)	160	≈16	0
2(HECC)	80	≈25	1
3(HECC)	54	≈76	1
4(HECC)	40	≈164	2

The smaller the objects are, the bigger their values are. Implementing Elliptic Curve Cryptography in many such devices helps in achieving efficient functioning of the state-of-the-art cryptographic algorithms. Elliptic Curve Cryptography implementation proceeds through a wide variety of domains.

A. Digital Post Marks

Innovative solutions and convenience seem to rule the age we are currently living in. Digital Post Marks (DPM) is one such technology initiative for an improved security for authorization and evaluation of products and services. DPM is the new look in digital postage that applies trusted time stamps by postal authority to ensure payments to the post office and validate digital signatures. The electronic document is locally signed by an individual, which is then authenticated by the Digital Postmark Service. If the signature is valid, a time stamp is generated as a counter signature. Digital Postmarking should be small and allow the post office to efficiently verify and detect fraud, thereby, mandating ECC as the perfect choice.

B. RFID Systems

Presently, the most important practical application in which light weight cryptography is needed is in automated identification of products. Radio Frequency Identification (RFID) is one such recent and advanced technology using radio frequency for identification. An RFID system consists of an antenna which reads the radio signal and transfers the information to a processing device called RFID tag. The RFID tag is an integrated circuit containing the RF circuitry and information to be transmitted. The RFID system can be incorporated just about anywhere, from products to missiles to pet tags, anywhere where a unique identification system is required. The major areas seeing applications of RFID tags are transportation, distribution and security.

Transport/Toll Payments— RFID operated cards have been successful as transport cards in place of tickets and often for several modes of transport. These RFID passes can be charged at metro stations and local convenience stores, and can be used in Metro, buses and parking lots. RFID tags are also used for electronic toll collection at toll booths. The variable tolls imposed are regulated by an active tagging system combined with the use of stored-value cards.

E-Passport— RFID technology is now being used in electronic passports issued by many countries. Along with biographic information, the electronic passport contains biometric information like digital photo or finger print recognition, stored on an RFID tag. With the updated features

to a regular passport, i.e., complex watermarks and a chip antenna, the biometric passports are counter-fraud initiatives for authenticating citizenship of travellers.

Automotive Controls— Microwave RFID tags are used in long range access control for vehicles. More automakers are offering high-tech, key-free systems called 'smart key' feature that let drivers enter and start their vehicles without a key. The key uses an active RFID circuit allowing the car to detect the key in the driver's pocket, briefcase or purse. The driver needn't remove it to wake up the engine.

Supply Chain Management— Many product vendors are now showing wide spread interest in placing RFID tags on all their shipments to speed up the distribution process and reduce the complexity of stock taking. RFID systems are also envisioned as replacing the older barcode technology of product identification. Barcode systems are simple to use although the amount of data that can be stored is limited. Another key advantage is that RFID scanning does not require line of sight and can be done at greater distances than bar code scanning. The replacement of bar-codes with RFID tags are also aimed at anti-counterfeiting of products..

C. Open SSL - Internet Security Protocol

Secure Communication is an intrinsic requirement for today's sensitive Electronic Data Interchange (EDI). EDI is a highly structured technique of conveying financial or personal information by the way of electronic commerce, web banking, electronic wallet or electronic mail. The latest version of the OpenSSL code containing ECC based cipher suites alleviates the computational burden on secure web servers due to their smaller key sizes and enables secure TLS/SSL handshakes.

D. Suite B

The National Security Agency (NSA), to complement its existing security, has announced a set of cryptographic algorithms called Suite B [18]. The suite which includes algorithms implementing Elliptic Curve Cryptography for digital signatures and key agreement is intended to provide the industry with a common set of cryptographic algorithms. They can be used to create products that meet the widest range of US Government (USG) needs.

It is no wonder that countries like the United States of America, the United Kingdom, Canada and some NATO nations have all adopted some form of elliptic curve cryptography for future systems to protect classified information throughout and between their governments. The Cryptographic Modernization Initiative in the US Department of Defence aims at replacing almost 1.3 million existing equipments over the next 10 years. In addition, the department's Global Information Grid will require a vast expansion of the number of security devices in use throughout the US Military.

This will necessitate change and rollover of equipment with all major US allies. Most of these needs will be satisfied with a new generation of cryptographic equipment that uses elliptic curve cryptography for key management and digital signatures ensuring added security.

V. Conclusion

The paper presented Light Weight Cryptography as the new generation public cryptography scheme for the future digital world. Compared to traditional cryptosystems like RSA, LWC using elliptic curves offers equivalent or even a better security with smaller key sizes, which results in faster computations; lower power consumption, as well as memory and bandwidth savings.

References

[1] *Elliptic curve cryptosystems.* **Koblitz, Neil.** 1987, Mathematics of Computation, Vol. 48, pp. 203-209.

[2] *A Normal Form for Elliptic Curves.* **Edwards, Harold M.** 2007, Bull. American Mathematical Society, Vol. 44, pp. 393-422.

[3] *Hyperelliptic cryptosystems.* **Koblitz, Neil.** 1989, Journal of Cryptology, Vol. 1, pp. 139-150.

[4] *Computing in Jacobian of a Hyperelliptic Curve.* **Cantor, D.** 177, January 1987, Math. Computation, Vol. 48, pp. 95-101.

[5] **Gaudry, Pierrick.** Genus 2 formulae based on Theta functions and their implementation. *11th Workshop on Elliptic Curve Cryptography.* [Online] September 5, 2007. http://mathsci.ucd.ie/~gmg/ECC2007Talks/ecc07-gaudry2.pdf.

[6] **National Institute of Standards and Technology.** Digital Signature Standard. *U.S. DEPARTMENT OF COMMERCE/National Institute of Standards and Technology.* [Online] January 2000. http://csrc.nist.gov/publications/fips/fips186-2/fips186-2-change1.pdf.

[7] **Certicom.** Certicom ECC Certificate Utility datasheet. *Certicom : Encryption and Security Document Library.* [Online] http://certicom.ca/download/aid-626/certutility_datasheet.pdf.

[8] *Comparing Elliptic Curve Cryptography and RSA on 8-bit CPUs - CHES 2004.* **Gura, Nils, et al.** Cambridge, MA, USA : Springer, 2004. Lecture Notes in Computer Science. Vol. 3156, pp. 119-132. CHES 2k4.

[9] **Lopez, J and Dahab, R.** *Performace of Elliptic Curve Cryptosystems.* 2000. Technical.

[10] **Kumar, Sandeep and Paar, C.** Are Standards Compliant Elliptic Curve Cryptosystems feasible on RFID. *Wokshop On RFID Security.* Graz, Austria : s.n., July 2006.

[11] *PRESENT: An Ultra-Lightweight Block Cipher.* **Bogdanov, A., et al.** s.l. : Springer Verlag Heidelberg Berlin, 2007. Cryptographic Hardware and Embedded Systems - CHES 2007. Vol. 4727, pp. 450-466. Lecture Notes in Computer Science.

[12] **Tuyls, P. and Batina, L.** RFID Tags for Anti-Counterfeiting. [ed.] In D. Pointcheval. *Lecture Notes in Computer Science.* February 2006.

[13] **Wolkerstofer, J.** Is Elliptic Curve Cryptography Suitable toSecure RFID Tags? *Workshop on RFID and Lightweight Crypto.* Graz, Austria : s.n., July 14, 2005.

[14] **National Security Agency, USA.** The Case for Elliptic Curve Cryptography. *National Security Agency, USA.* [Online] http://www.nsa.gov/ia/industry/crypto_elliptic_curve.cfm.

[15] **Pelzl, Jan, et al.** Hyperelliptic Curve Cryptosystems: Closing the Performance Gap to Elliptic Curves (Update. *Cryptology ePrint Archive.* [http://eprint.iacr.org/]. March 2003.

[16] **Bernstein, Daniel J.** Elliptic vs. hyperelliptic, part 1. *cr.yp.to.* [Online] September 20, 1006. http://cr.yp.to/talks/2006.09.20/slides.pdf.

[17] **Bernstein, Daniel J.** Elliptic vs. hyperelliptic, part 3 : Elliptic strikes back. *cr.yp.to.* [Online] May 22, 2007. http://cr.yp.to/talks/2007.05.22/slides.pdf.

[18] **National Security Agency, USA.** Fact Sheet NSA Suite B Cryptography. *National Security Agency, USA.* [Online] http://www.nsa.gov/ia/industry/crypto_suite_b.cfm.

[19] *Comparing Elliptic Curve Cryptography and RSA on 8-bit CPUs.* **Gura, Nils, et al.** s.l. : Springer Verlag Heidelberg Berlin, 2004. Cryptographic Hardware and Embedded Systems - CHES 2007. Vol. 3156, pp. 119-132.

[20] **Batina, L., et al.** An Elliptic Curve Processor Suitable For RFID-Tags. *Cryptology ePrint Archive.* 2006. http://eprint.iacr.org/.

Energy Dependent Connection Availability Model for Ad Hoc Networks

Dimitar Trajanov, Sasho Gramatikov, Sonja Filiposka and Aksenti Grnarov

Dept. of Computer Sciences
Faculty of Electrical Engineering and Information Technology
University Ss. Cyril and Methodious Skopje
Skopje, R. Macedonia
mite@feit.ukim.edu.mk
saso.gramatikov@feit.ukim.edu.mk
filipos@feit.ukim.edu.mk
grnarov@feit.ukim.edu.mk

Abstract – **Although wireless networks have existed for many years already, explicit concern about their energy efficient operation has emerged only recently with the proliferation of lightweight mobile devices. In this paper we propose an energy dependent connection availability model for ad hoc networks which relies on a number of real measurable parameters. The model provides the means to analyze and optimize the energy consumption in the ad hoc network while taking into consideration the nodes connectivity and the limited battery capacity with possibilities for recharging**.

Index Terms **– Energy Consumption, Ad Hoc Networks, Connection Availability, Battery Recharging.**

I. Introduction

Each day mobile devices are getting smaller in size, lighter in weight and more powerful in processing power. However the biggest weakness of this fashion is the limited power supply which imposes a strong impact on the environment independence. This feature of the ad hoc networks is achieved by using rechargeable batteries. In order to produce batteries with acceptable size and weight for a mobile device, it is necessary to either increase the energy density of the battery, or to reduce the energy consumption by the mobile device. Although nowadays the battery capacity has considerably increased, it can not keep up with the raising necessity of functionality and diversity of the mobile devices.

Wireless communication is considered one of the standard features that every mobile device should posses. Although the most widely spread IEEE 802.11 standards have taken into account the need for energy efficiency, the reducing of energy consumption still remains a major issue. The necessity of being online and able to interchange large quantities of data like speech, pictures and videos leads to a constant activity of the wireless interface. This corresponds to a situation where the energy consumed by the wireless interface takes a significant part of the overall energy consumption of the wireless device[1]. Such an example are the notebooks which in average consume 5 to 15 W, while the active wireless interface consumes approximately 1.5 W, which makes 15 to 35 % of the overall energy consumption [7]. This property is even more emphasized in PDA devices where the energy consumption of the processor, memory and screen is reduced to minimum.

Reducing the energy consumption of the wireless interface is very complex since a big amount of the consumed energy is being wasted for emission of electromagnetic waves with a certain power necessary to reach the destination in order to have communication establishment. One of the problems that should be taken into consideration when reducing the energy consumption of the wireless interface, is not to allow the connectivity to become an issue of concern. Therefore, in this paper, we make an analysis of the energy consumption in ad hoc networks in order to study the influence of the energy efficiency on the ad hoc network connectivity and availability.

The previous work can be mainly divided in two groups. The first one considers models of energy consumption by the wireless interface, whereas the other group considers models of connection availability in ad hoc networks.

The proposed models of energy consumption by the wireless interface takes into consideration many parameters, including energy consumed by the transmitter electronics, the minimum energy required for achieving a desired bit error rate, the environment dependence etc. L. Feeney in [1][5] presents a model for evaluating the energy consumption of wireless interface and it is used to examine the energy consumption of AODV and DSR ad hoc routing protocols. In [7] are given measurements of the energy consumed by the wireless interface with varying the data packet size, transmission rate and RF power level. In [7], starting from a detailed radio model, they create a model of energy consumption by the wireless interface. This model is used for finding the optimum one-hop transmission distance that will minimize the total system energy. In [2] an analytical model to investigate the optimal value of the radio transmission range, based on the underlying device energy consumption model and a two-dimensional Poisson node distribution, is proposed. In [4], based on the simplified energy model, they create model for radio range adjustment that will result in optimizing energy consumption in wireless sensor networks. In [8] there is an overview of energy concerns in wireless networks and a

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 489–494.

simplified model for energy consumption by the wireless interface.

The second group, models ad hoc network connection availability depending on node failures. In [11] the continuous time Markov chain (CTMC) is used to represent a connection availability model for a two hop ad hoc network that incorporates the physical faults like: node, power and link faults. In [12] a generalized model with enlarged number of node failure types is introduced. These models give acceptable results only for static ad hoc networks. In mobile ad hoc networks, the probability that a connection will be broken as a result of the routing node mobility is far greater than the probability that a physical fault will occur at the routing node. In [10] generalized connection availability model based on real measurable parameters that concern the performances of mobile ad hoc networks is proposed.

In this paper we propose model where we take into consideration the traffic intensity generated by each node and the energy consumed by the rest of the components of the wireless device as highly dependent parameters on energy consumption. We implement this model into [10] in order to get an improved ad hoc connection availability model, where among the others parameters traffic and power consumed by other components, will have significant role.

The rest of this paper is organized as follows. In section II the battery model is presented. Section III covers the energy consumption model and derivation of the energy dependent node availability. Section IV expands the model for cases when battery recharging is feasible. In the next section, the influence of energy consumption on ad hoc network connection availability is presented. The results of the analysis of influence of the energy consumption model for ad hoc network are given in section VI after which follows the conclusion and some remarks about possible future work in this field.

II. THE BATTERY MODEL

There are a number of versatile technologies of rechargeable battery production on the market which are used as power supply in the mobile devices. The examples include Nickel Cadmium, Nickel Metal Hydride, Lithium Ion, Lithium Polymer, Lead Acid and Alkaline. From all of these, the Lithium Ion gives the best performances i.e. they offer the highest energy density in the battery packs.

The parameters which are of interest in the battery production are energy density measured in Wh/kg with values from 30 to 150 Wh/kg, cycles for reaching 80% of the total capacity with values from 50 to 500, recharge duration with values from 1 to 4 hours and the percentage of monthly self discharging which can reach values up to 30 % [6].

All electrical cells are characterized by an initial open circuit potential which is the value of the potential of a fully charged cell under no load-conditions. The cut off potential is the minimum voltage achieved until the battery is considered discharged. There are two types of capacities which characterize the batteries: a theoretical and nominal capacity expressed in Ampere hours (Ah). The theoretical capacity is based on the total amount of active materials in the cell and represents the maximum available energy of the cell. This is never achieved in practice and therefore when talking of capacity we consider the nominal capacity which is the energy obtained when the cell is being discharged with a constant current until the cut off potential is reached. The nominal capacity is always lower then the theoretical.

III ENERGY CONSUMPTION MODEL

When modeling the energy consumption in wireless communication, one significant feature that should be taken into consideration is the radio wave propagation fading, defined as energy loss due to reflection, diffraction and scattering [9]. Therefore, the process of choosing the minimal level of transmission power for each node in order to establish a communication with the other nodes in the network is very complex and is typically done in statistical fashion, based on a great number of measurements. Very often there are situations where for a given transmission power, two nodes can communicate in one direction only, whereas the connection can not be established in the opposite direction due to obstacles [1].

Based on models proposed in [2] [3] [4] and [8], which are taken to give satisfying results, we use a simplified model of the energy consumption when sending and receiving a single bit. When a single bit of information is sent through a wireless environment, the energy consumed by the network interface can be obtained by the relation:

$$E_{Sbit}(r) = E_t + E_d r^{\omega} \quad (1)$$

where E_t is energy consumed by the transmitter electronics and $E_d r^{\omega}$ is the minimum energy required for achieving a desired bit error rate. E_d represents the energy of the signal just before it is being sent to the receiver and $r^{-\omega}$ is the attenuation due to propagation fading in the environment. ω is the attenuation factor and it depends on the properties of the wireless environment having values from 2 to 4.

The energy consumed when receiving one bit is constant and it is given as [4]:

$$E_{Rbit} = E_r \quad (2)$$

where E_r is the energy spent by the receivers electronics when receiving one bit of information. It should be noted that the energy consumption per bit, depends on the bit rate of the transmission when sending or receiving data.

Nodes in an ad hoc network can act as both hosts and routers since they can generate and forward packets. The overall energy spent by a node for sending and receiving one bit will be the sum of the energy consumed for sending and the energy consumed for receiving a bit, multiplied by the probabilities that a single bit is being sent α or received β,

$$E_{bit} = \alpha E_{Sbit} + \beta E_{Rbit} = \alpha E_t + \alpha E_d r^{\omega} + \beta E_r \quad (3)$$

In order to simplify the expressions, we analyze the case when the station sends and receives a bit with equal probability $\alpha = \beta = 0.5$.

We are considering a mobile device where the overall energy consumption is divided into two parts: energy consumed when sending or receiving data packets via the wireless network interface, while the second part includes the total energy consumed by all of the components of the mobile device (processor, screen, hard drive, memory, idle state of wireless interface).

Let us assume that the device has a constant power dissipation when in working state, P_{device} . Thus, the battery capacity C will be:

$$C = kBE_{bit} + tP_{device} \tag{4}$$

where k is the number of packets generated within the (0, t) interval and B is the average packet length in bytes (including the packet header and the routing protocol packets which are being sent in order to establishing the end-to-end communication).

The maximum number of packets that will be generated within the battery life time T is

$$K = \left\lfloor \frac{C - TP_{device}}{BE_{bit}} \right\rfloor \tag{5}$$

If we assume that the source node generates packets according to the Poisson distribution, we can model our ad hoc station like a queue with arrival rate λ and service time with a Poisson distribution. The probability that a node will be active in a certain moment is equal to the probability that less than the maximum of K packets will be generated until the end of the maximum battery lifetime. According to the Poisson distribution, the probability that exactly k packets are generated within the (0, t) period is:

$$P(k,t) = \frac{(\lambda t)^k}{k!} e^{\lambda t} \tag{6}$$

The probability that less then K packets are generated within the time interval T is a sum of the single probabilities that exactly 0, 1, 2 … *K-1* packets are generated:

$$P_{active}(t,K) = P(k < K,t) = \sum_{n=0}^{K-1} \frac{(\lambda t)^n}{n!} e^{\lambda t}$$

$$P_{active}(t,K) = \frac{Gamma(K,\lambda t)}{Gamma(K)} \tag{7}$$

where $Gamma(K)$ is the Euler's Gamma function and $Gamma(K,\lambda t)$ is the incomplete Gamma function.

The distribution function $P_{active}(t,k)$ also represents the distribution of the node reliability $R(t,k)$. Thus, the Mean time to failure MTTF can be obtained as:

$$MTTF = \int_0^\infty R(t,k)dt = \int_0^\infty P_{active}(t,k)dt \tag{8}$$

Since the MTTF can not be presented in analytical closed form solution, (k is dependent on time in the integral), we used numerical methods in order to obtain its value.

IV Node Availability With Battery Recharging

Since the mobile nodes use a battery power supply, then once the battery is discharged it is necessary to recharge or change it with a new one. The recharging or changing operation disables the functionality of the mobile node in the ad hoc network for a certain time (it is assumed that the power supply is not immediately available anywhere in the network area). The amount of time the node is not able to be part of the network is variable (whether the battery is being recharged or changed).

When considering this scenario, we model the system in its stationary state wherein the probability that a certain node is active equals to the nodes availability. It is defined as the ratio between the mean time of failure of the node MTTF and the sum of MTTF and the time necessary for recharging the battery. The probability that a node is active is given with:

$$P_{active} = \frac{MTTF}{MTTF + t_c} \tag{9}$$

where t_c is the time spent in order to recharge or change the battery.

V Energy Dependent Connection Availability of Ad Hoc Network

One of the most important issues for mobile ad hoc networks is to know the availability of the system. In order to investigate influence of energy efficiency on connection availability of Ad Hoc Networks we use the Connection availability model proposed in [10]. The model is in closed analytical form and includes influence of large number of parameters like: routing protocol, number of participants in ad hoc network, distance between source and destination node, nodes speed, transmission range, and area size where participants of ad hoc network are scattered.

The model is based on a common simplified view of the ad hoc network where all the mobile nodes have the same transmission power and are equipped with omni directional antenna, thus having equal transmission range r. In the model N+2 nodes placed in area A are observed (see Fig. 1). The two additional nodes represent the source and the destination for the analyzed end-to-end connection. The N nodes play the role of routers in the connection path between the source and the destination. While moving around in A, a node can enter the B

area and, after a certain period of time, leave *B* and enter area *C* defined as *A*-*B*. This process is continuously repeated. While establishing the communication between the two nodes, MNs (source) and MNd (destination) the communication path has to go through one of the nodes that are currently located in the intersection area *B* between MNs and MNd (i.e. MN1, MN2). In order to obtain a closed form solution for the connection availability, a system model solvable by a Continuous Time Markov Chain (CTMC) is used.

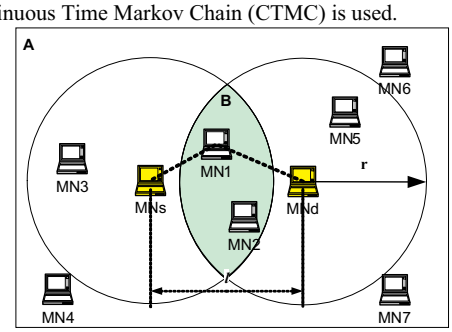


Fig. 1 Ad hoc network model

The steady state connection availability (SSCA) is given by:

$$A_s = \sum_{k=1}^{N} \frac{N!}{k!(N-k)!} \left(\frac{\mu}{\eta}\right)^k \left(\frac{\delta_r}{\eta+\delta_r}\right) \pi_{0,3}$$

$$\pi_{0,3} = \frac{(\eta+\delta_r)/\delta_r}{\left(1+\frac{\lambda}{\delta}\right)\left(1+\frac{\mu}{\eta}\right)^N + \eta N \mu \left(\frac{1}{\delta_r}-\frac{1}{\delta}\right)} \qquad (10)$$

where η is the rate of leaving the intersection region *B* and depends on the distance between the nodes, transmission radius, size of area where nodes are scattered (area *A*) and nodes speed; μ is rate of returning in *B,* and depends on the same parameters like η; *N* is number of network participants (nodes); The influence of routing protocols is introduced with two parameters: average switching delay $1/\delta$ and connection reestablishment delay $1/\delta_r$. When there is at least one node in the intersection region the average time needed for connection switching is defined as average switching delay $1/\delta$. Otherwise, the average time needed for connection reestablishment is defined with the average connection reestablishment delay $1/\delta_r$. When using the AODV routing protocol, the average switching delay is taken to be 0.0565s whereas the connection reestablishment delay is dependent on the amount of elapsed time and the leaving rate. More details are presented in [10].

In this model for connection availability, one of the parameters of dependence is the mean number of stations which is considered to be invariant in time. The performances and failure tolerance of the ad hoc network depend on a large scale of the number of stations that take part in the communication. Because of their mobility, the only possible power supply of the ad hoc nodes is the battery, which imposes that after a certain time of activity, the battery of the node will discharge. Once the battery is completely discharged those nodes are cut off of the network. Consequently, the number of nodes in the ad hoc network will constantly change. Therefore, it is necessary that we define an energy model that will illustrate the change of the number of nodes in the network caused by the energy parameters.

Each node in the network, under certain conditions (moment in time, battery status, traffic intensity…) is active with probability P_a, such that its activity is independent on the other nodes. Hence, the probability that *m* out of *N* total nodes are active is given with the binomial distribution

$$P_a(m) = \frac{N!}{m!(N-m)!} \left(P_{active}\right)^m \left(1-P_{active}\right)^{N-m} \qquad (11)$$

The mean number of active nodes is given by

$$\overline{N}(t) = NP_{active}(t) \qquad (12)$$

Substituting the expression for the mean number of nodes (12) into SSCA in (10) we acquire an expression which includes the influence of the energy related parameters on the overall nodes availability of the mobile ad hoc network.

VI Analysis of the Energy Consumption Influence on Ad Hoc Network Availability

In order to analyze the influence of the energy consumption on the nodes availability, we have to take into consideration the commonly used values for the wireless interface parameters. We use the following parameters according to [5]: energy per bit spent by the transmitter's electronics $E_t = 50 \cdot 10^{-9} J/bit$, energy per bit consumed by the amplifier when preparing the signal for transmission $E_d = 100 \cdot 10^{-12} J/bit/m^2$ and energy per bit spent by the receiver . These values refer to a bit rate of $1\,Mbps$. The path loss parameter ω is set to 2 referring to an even environment without obstacles. The transmission range is set to 250 m. The packet size is 512 Bytes, the battery capacity is 3000 mAh, device power is 3 W and the packet are sent with rate of 10 packets/s.

A. Single Node Availability Analysis

If it is not possible to change or recharge the battery of the ad hoc network devices, than the probability that each node is active depends on the elapsed time. One of the crucial parameters that affect the reliability of the node is the transmission radius because the energy consumption is heavily dependent on the power level needed to transmit the signal over farther distances.

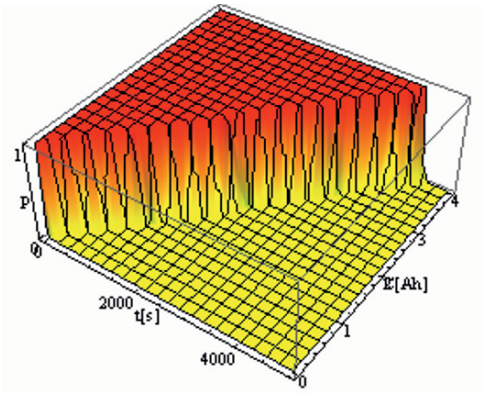


Fig. 2 Node availability depending on the battery capacity

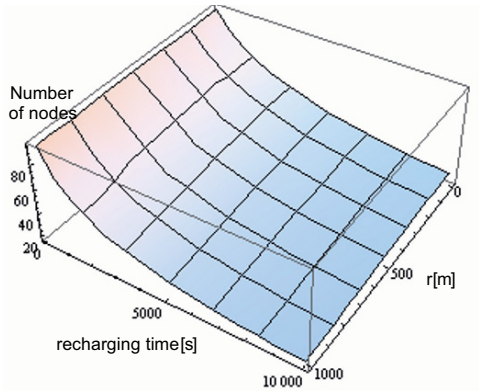


Fig. 4 Node availability depending on the recovery time and transmission range

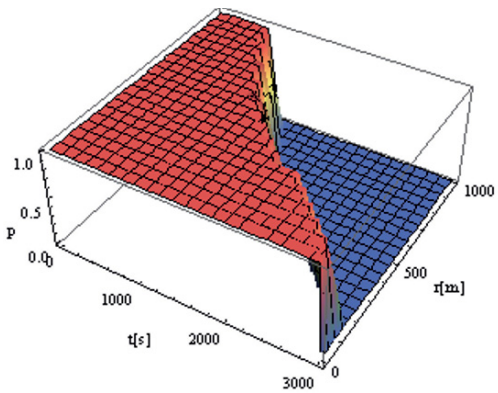


Fig. 3 Node availability depending on the transmission radius

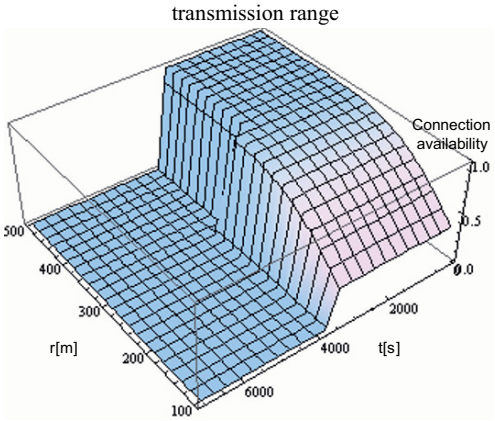


Fig. 5 Node connection availability depending on the transmission range without possibility to recharge/change the battery

On Fig. 2 the linear dependence of the node availability on the battery capacity is presented. High availability can be achieved by increasing the battery capacity, but at the same time we must keep in mind that the capacity can not be increases infinitely as it also increases the size and price of the mobile device.

On Fig. 3, the node availability depending on the transmission radius is shown. The probability that a node is active equals one, and reduces rapidly to zero, as the battery completely discharges. For larger values of transmission radius, the energy consumption rises because of the need for more powerful signal due to the increased propagation fading.

Similar behavior can be observed when varying the packet size parameters or the device power dissipation. The battery would last longer for smaller packets of data whereas it discharges rapidly for larger data packets.

Fig. 4 presents the dependence of the node availability on the recharge time t_c and the transmission range. The changes are very smooth for any deviation of the depending parameters, reaching lower values for higher values of the recovery time and transmission range.

B. Connection Availability Analysis

On Fig. 5 the dependence of node connection availability on the transmission radius and elapsed time is presented without possibility to recharge/change the battery. The total number of nodes is 100 and they move within a 1000 m x 1000 m area with average speed of 1 m/s according to the Random Direction mobility model. The distance between the source and destination node is 1.5 x transmission radius. It is evident that the availability is relatively constant with time, and after a certain period of time it rapidly falls to zero value. This is as a result of complete battery discharging and the reduction of the number of nodes in the network. For smaller values of the transmission range the connection availability is very small because of the lack of connectivity between the network nodes.

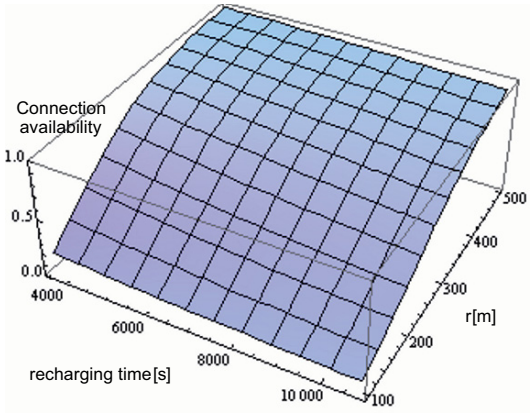


Fig. 6 Node connection availability depending on the transmission range with possibility to recharge/change the battery

On Fig. 6 the dependence of node connection availability on the transmission radius and recharge/change time is presented. The availability is relatively constant and slightly reduces when the recharge/change time increases. In this case, the small transmission range also heavily decreases the connection availability.

VII. Conclusion

The energy consumption is one of the most important issues related to the availability of nodes in ad hoc networks. In order to make the nodes life last longer, the energy consumed during its operation should be reduced to a minimum.

In this paper an energy dependent connection availability model for ad hoc networks is presented. The model is based on a number of measurable real network energy dependent parameters like: battery capacity, unit energy consumption for sending and receiving packets, traffic size, power consumption of the mobile device, recharge time; and protocol and spatial dependent parameters like: transmission range, routing protocol, number of participants in ad hoc network, distance between source and destination node, nodes speed, and area size where participants of ad hoc network are scattered.

The model can be used in order to make a detailed analysis of the influence of the energy consumption on the connection availability in the ad hoc network. Thus, the amount of influence of each of the separate parameters can be obtained which can afterwards lead to modifications of the most critical parameters with the purpose of achieving higher energy efficiency.

One of the most important advantages of the proposed model is that it is possible to make an optimization of the energy consumption while sustaining the connection availability in the ad hoc network. In this way we are certain to obtain an ad hoc network with low power consumption and, in the same time, desired connection availability.

Generally we can conclude that when there is no possibility of recharging the battery, the transmission range hardly affects the connection availability, whereas when battery recharging is possible, the transmission range has a considerable influence of connection availability.

VIII. Future Work

One of the simplifications we made in the creation of our energy model of an ad hoc network was the constant power consumed by the components of the mobile device. In order to have a more realistic energy model, there should be a separate energy model of the components of the device that are major energy consumers, like the processor, hard drive, screen and memory. Each of these models would be incorporated in the representation of the device power consumption in our model. Other complexity that can be added in the model is the idle state of the wireless interface and the different probabilities of being in each of the send receive and idle state, as well as the probabilities for the wireless activity of the other nodes in the network. Having a model where the activity and energy consumption of every component of a mobile device is included, could serve for optimizing the operation of each component for minimum energy consumption.

References

[1] Laura Marie Feeney, “Energy Efficient Communication in Ad hoc Wireless Networks” Chapter ENERGY EFFICIENT COMMUNICATION in Stefano Basagni, Marco Conti, Silvia Giordano, and Ivan Stojmenovic, editors, Mobile Ad Hoc Networking. Wiley-IEEE Press, July 2004.

[2] Jing Deng, Yunghsiang S. Han, Po-Ning Chen, Pramod K. Varshney, “Optimum Transmission Range for Wireless Ad hoc Networks”, 3rd IEEE Upstate New York Workshop on Sensor Networks, New York, USA, 2004.

[3] P. Chen, B. O’Dea, and E. Callaway, “Energy efficient system design with optimum transmission range for wireless ad hoc networks”, IEEE ICC, pp. 945–952, 2002.

[4] Q. Gao, K. J. Blow, D. J. Holding,I. W. Marshall, X. H. Peng,,” Radio Range Adjustment for Energy Efficient Wireless Sensor Networks”, Ad-Hoc Networks Vol. 4, No. 1, Jan 2006, pp 75 - 82. ISSN: 1570-8705

[5] Laura Marie Feeney, “An Energy Consumption Model for Performance Analysis of Routing Protocols for Mobile Ad Hoc Networks”. Mobile Networks and Applications 6, 239–249, 2001

[6] “Computer Desktop Encyclopedia”, Computer Language Company Inc, (www.cadex.com), 2005.

[7] Jean-Pierre Ebert, “Energy-efficient Communication in ad hoc Wireless Local Area Networks”, PhD Dissertation, Fakultät Elektrotechnik und Informatik der Technischen Universität Berlin, 2004.

[8] Anthony Ephremides, “Energy concerns in wireless networks”, IEEE Wireless Communications, August 2002

[9] Theodore S.Rappaport, “Wireless communications”, Prentice Hall PTR communications engineering and emerging techonologies series, 2002

[10] Trajanov Dimitar, Filiposka Sonja, Efnusheva Marija, Grnarov Aksenti, “Ad Hoc Networks Connection Availability Modelling”, *ACM PE-WASUN*, Italy, 2004

[11] Chen D., Garg S., Trivedi S. K., “Network Survivability Performance Evaluation: A Quantitative Approach with Applications in Wireless Ad-hoc Networks”, MSWiM ‘02, 2002, Atlanta, USA

[12] Koroma J., Li W., Kazakos D., “A Generalized Model for Network Survivability”, TAPIA’03, Atlanta, USA, 2003

Trust Management in Ad Hoc Network for Secure DSR Routing

Subhrabrata Choudhury, Suman Deb Roy, Sneha Aman Singh
National Institute Of Technology, Durgapur
Durgapur, India
subhrabrata@gmail.com, suman_debroy@yahoo.com, amansingh.sneha@gmail.com

***Abstract* - An ad hoc network is wireless, self-configuring network of mobile routers forming a dynamic topology. In such an environment, any malicious behavior by these routers (nodes) could greatly compromise secured routing. However, just detecting the malicious nodes is not enough to guarantee security. It should be ensured that, only trusted nodes take part in routing. In this paper, we propose a distributed dynamic trust evaluation scheme for securing DSR routing in mobile ad hoc and sensor networks. Each node periodically calculates trust for others based on its direct experience, recommendation and previous knowledge. Extensive simulation studies show that our scheme, unlike most of the previously proposed ones, efficiently adapt to fast changing trust scenario and the trust converges to values, revealed by long term behavior of the nodes. This scheme is resistant to 'collusion' attack and is cautious not to punish nodes when packet is dropped due to 'node mobility'.**

Keywords- ad hoc network, trust management, DSR, secured routing, collusion attack.

I. INTRODUCTION

In wired network, routers are secured and we trust our Internet Service Providers (ISP). ISPs trust each other and send packets from source to destination through a series trusted routers. With the help of Border Gateway Routing Protocol (BGP) [19] and IP Security (IPsec) [20], the ISPs can guarantee a trusted transmission throughout the Internet. Thus, though the application maybe vulnerable to malicious attacks yet the routing services throughout the network remain relatively secured.

Ad Hoc Network is a "temporarily formed wireless network which are created, operated and managed by the nodes" as defined in [3] (mobile hosts with wireless network interfaces) themselves. The operation and survival of an ad-hoc network is solely dependent upon the cooperative and trusting nature of participating nodes. But unfortunately, the absence of fixed infrastructure such as a certifying authority (CA) makes it impossible to establish trust relationships based on the conventional authentication protocols.

Routing protocols, in general, assume that the nodes are trustworthy and co-operative. This dependency of routing protocols on the behavior of intermediate nodes makes the ad-hoc network vulnerable to passive and active attacks by malicious nodes [4]. The malicious nodes aim to eavesdrop on, corrupt, or disrupt the network traffic. In Passive attacks [3] the attacker does not disturb the routing protocol. It only eavesdrops on the routing traffic and endeavors to extract valuable information like node hierarchy and network topology from it without letting other nodes know about the attack. In active attacks [3], the attacking node has to spend some of its energy to carry out the attack.

In this paper, we propose a Dynamic Trust Establishment Scheme (DTES) based on 'vectored trust model' [2] in the context of DSR [1] routing in a purely ad hoc environment. For this purpose, 'trust of some degree' is represented as weighted sum of individual components viz. – 'experience', 'recommendation' and 'knowledge'. Each node maintains and updates its trust values for all other nodes in the network in the form of a trust table. The experience component of trust for each neighbor is directly measured by the nodes for its immediate neighbors and kept updated at regular intervals. The existing trust table is broadcasted to all other nodes. Trust-table received from other nodes is used to calculate the recommendation part of the trust. At a regular interval, the previously evaluated trust is included in the current knowledge component of total trust.

We organize our paper as follows: In Section II, reviews of previous work is provided along with brief discussion on vectored trust model. Our scheme is discussed in detail in Section III. The simulation scenario and experimental results are discussed in Section IV. In Section V we conclude our paper with outline of the future scope.

II. PREVIOUS WORK AND VECTOR TRUST

A. Previous Work

Considerable amount of work has been done in the field of trust management. In [6], Balakrishnan suggests trust management using cryptographic mechanisms. Cryptographic techniques often constitute complex encryption/decryption algorithms and are not very suitable for ad hoc environment and less so for sensors. Simplicity is a critical factor in such systems as has been analyzed in [5]. Pirzada and McDonald [3] seek to establish trust in pure ad hoc networks using the concept of weights. However, they neglect the combination of third basic entity that constitutes trust in any society namely – '*knowledge*' which learns from the past incidents. Catharina and Cary in [7] choose to compute incomplete trust; where trust is calculated for a third node using a second node

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 495–500.

hierarchically even if the calculating node has no experience with the third. Such a model could be prone to random attacks if the second (recommending node) is itself malicious. Reference [8] and [9] stands on the idea of global trust model alone which aggregates the recommendation component without considering its authenticity. In the absence of local trust, a node receiving recommendation will not be able to judge the trustworthiness of the recommender. We choose DSR as our routing protocol in spite of AODV's effectiveness since in constrained resource situations, very common to an ad hoc environment, DSR comprehensively outperforms AODV [10].

Our DTES is proactive whereas [16] is based only on pre-established routes. We use weightage to remove absolute considerations so that sufficient importance is given to crucial parameters. This is a vital requirement for any changing environment and is avoided in many previous works.

B. Vector Trust

Here we outline the vectored trust model as proposed in [2], where trust is represented as the weighted sum of three components, viz – knowledge, experience, and recommendation

$$ {}_x{}^c\mathrm{T}^y = W_k * {}_x{}^c\mathrm{T}^y_k + W_e * {}_x{}^c\mathrm{T}^y_e + W_r * {}_x{}^c\mathrm{T}^y_r \quad (1) $$

where, ${}_x{}^c\mathrm{T}^y$ is the trust of x on y in the context c, ${}_x{}^c\mathrm{T}^y_k$ is the value of knowledge component of trust of x for y in the context c, ${}_x{}^c\mathrm{T}^y_e$ is the value of experience component of trust of x for y in the context c, ${}_x{}^c\mathrm{T}^y_r$ is the value of recommendation component of trust of x for y in the context c, W_k, W_e, W_r are the weight of knowledge, experience and recommendation component respectively.

Trust values allotted to the nodes are based on the following components:

- Experience - It is evaluated by a node x for its neighbor y in terms of the number of successful and unsuccessful interactions that has occurred between them in the context c within a specified period of time $[t_n, t_{n+k}]$ where n is a whole number.
- Recommendation - It is evaluated by a node x for the node y of the same network on the basis of recommendation value (trust value) as returned by each of the other nodes on that network acting as its recommender. Node x uses the "level of trust" he has on the recommender in the context "to provide a recommendation" as a weight to the value returned. This weight multiplied by the former value gives the actual recommendation value for trustee y by x in the context c.
- Knowledge – Knowledge provides information about the previous interactions of a node with its peers in the ad hoc environment. This parameter is generated from a combination of the present experience, present recommendation and the past (previous) knowledge itself. This means knowledge is such an entity which grows on itself, in addition to some external inputs every round. In other words, past knowledge is utilized to calculate present knowledge (and hence trust).

III. Our Approach

In our paper, dependence on a central trust authority is considered to be an impractical requirement, as pure ad-hoc network follows a changing topology. The model is built such that the mobility of node, hence the dropping of packets because of the node mobility, does not affect the trust value of corresponding related nodes. If no events take place in a node in the time interval $[t_n,t_{n+k}]$, then test packets are forwarded to the node to update the trust value.

A. Dynamic Trust Establishment Scheme (DTES) and Update Management

As complete trust and/or distrust and even discrete representation of trust cannot be assigned to a node in ad hoc network, trust is represented within continuous value ranging between 0 and 1, with 0.5 as the neutral value. The update management essentially is a subset of Trust Management System (TMS). We adapt the V-trust model outlined in the above section in the context of DSR routing in MANET based on the following steps:

- Initially trust value allotted for all the nodes of the network is +0.5. Thus, if there are 'N' numbers of nodes, then trust vector can be represented as:

  (0.5, 0.5, 0.5, …….., to N terms)

- The trust vector of each node is broadcasted to all other nodes at regular time interval t_k required to update the trust table.
- The trust of each node is updated on the basis of any of the following events:

  a) When acknowledgement ACK for original data packets comes back.

  b) When acknowledgement ACK for test packets comes back.

  c) When recommendation from any other node comes

- In case of positive behavior of the node in the network, the trust value of node is increased and in case of negative behavior of any node leading to events such as dropping of packets or misrouting a packet, the trust value of the node is reduced. This follows the generic trust allotment trend and is explained in Section III-B-1 with (2) and (3).
- Rate of decrease of trust value of a node in case of negative behavior is greater than rate of increase of trust value of that node in case of positive behavior. The practice is a direct abstraction from the characteristic of social networks (whether human or ad hoc), where earning respect from peers takes a number of good deeds stretched over a long time; whereas losing it takes only a few misdeeds. We integrate this natural trend in our approach by giving more

weightage to punish the malicious node than to award a good node. Thus, a node earns trust from its peers steadily but could lose it drastically if it engages in malicious behavior.

B. Trust Computation

A node x decides a trust value for node y, say, ${}^{x}T_{y}$ depending on the various situations or contexts which in turn is dependent on the protocol and the scenario to which the trust model is being applied.

1) Trust calculated on the basis of Experience: Experience is gained by two nodes only when they have interaction with each other. Hence, only neighbors of a particular node can compute the trust value of that node on the basis of experience. Let trust value computed on the basis of experience be represented as T_e. For simplicity experience regarding neighbors is directly measured.

Though experience about other nodes (not neighbors) can also be computed using graph theoretic techniques but here for the reason of simplicity, we do not include it. This is particularly important when the nodes are having less computing power such as sensors.

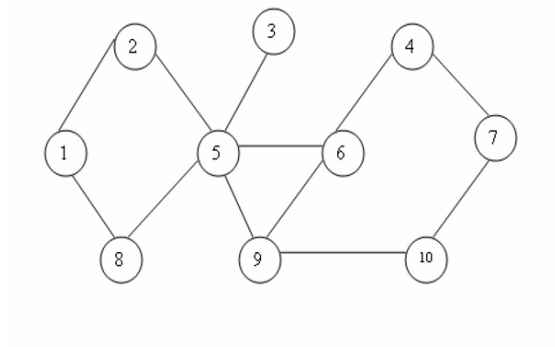


Figure 1. An instant snapshot of an ad hoc environment

For example, in Fig.1, node 1 can calculate trust value T_e of nodes 2 and 8 only, as they are the only neighbors of node 1. Other than these nodes, node 1 cannot directly interact with any other nodes.

Node 1 calculates the trust value T_e of nodes 2 and 8 by either of the two means:

a) If one or more than one event occurs in the time interval t_k in which node 1 sends data packets through either node 2, then T_e is computed for node 2, taking into account net successful interactions that has taken place.

b) If normal data packets are not available to be sent to any of the nodes 2 or node 8 within the time period t_k, then node 1 sends test packets through that node, and T_e is computed taking into account net successful interactions that has taken place.

Trust value for node 2 computed by node 1 on the basis of experience T_e can be given as:

When $S \geq U$,

$$T_e = 1 - \frac{1}{\left\{\left(\frac{S-U}{S+U}\right) * W_{is}\right\} + 2} \quad (2)$$

When $S < U$,

$$T_e = \frac{1}{\left\{\left(\frac{U-S}{S+U}\right) * W_{iu}\right\} + 2} \quad (3)$$

Here, S represents the number of successful interactions, U represents the number of unsuccessful interactions, W_{is} and W_{iu} represent weight of net successful interaction and unsuccessful interactions respectively which depend on the number of interactions taking place. The value of W_{is} and W_{iu} is more when number of interactions is high. W_{is} is always less than W_{iu} since in our model, rate of decrease of negative trust is more than the rate of increase of positive trust.

Equation (2) confirms the fact that if $S \geq U$, $T_e = [0.5, 1)$. So, a good node always gets an experience of 0.5 or more on the scale of 0-1. On the other hand, (3) represents cases where $U > S$, and hence $T_e = [0, 0.5)$ which shows that packet dropping nodes are given low trust values.

2) Trust calculated on the basis of Recommendation: A node x can decide over the trust value of node y by taking recommendations from other nodes which may or may not be neighbors of y depending on two factors:

a) The trust values of node y as computed by all other nodes in the network other than nodes x and y. This is basically the *recommendation* for y.

b) The knowledge of x about the trust value of the nodes that has broadcasted the trust value of node y, which was again obtained by the similar method in case the recommenders are not the direct neighbor of the node x which is calculating the trust of node y. This involves *checking the trust of the recommender itself*. It means weightage given to a particular recommendation depends on how much trusted the recommender itself is. We included this functionality to reduce the chances of a malicious node misreputing its peers. Since a malicious node itself has low trust, its recommendation (which could be intentionally crafted to degrade the system) will not be given much importance by the node receiving the respective recommendations.

Hence the recommendation value obtained by node x about node y from all other nodes of the network can be represented as:

$${}_{x}^{r}T^{y} = \sum_{i \neq y, x} \left({}_{x}T^{i} * {}_{i}T^{y} \right) / \left(\sum_{i \neq y, x} {}_{x}T^{i} \right) \quad (4)$$

where, ${}_xT^i$ is the trust of i^{th} node in the network as given by node x such that $i \neq y$ and $i \neq x$, ${}_iT^y$ is the trust value of the y as broadcasted by the i^{th} node in the network, that is, the recommender.

3) Trust calculated for the current time interval: After the trust value has been measured on the basis of the two parameters experience and recommendation for the time interval t_k, the total trust of node y calculated by node x for that time interval can be given as:

$$T_{tk} = W_e T_e + W_r T_r \quad (5)$$

such that,

$$W_e + W_r = 1 \quad (6)$$

for weights W_e and W_r. If node x is the direct neighbor of node y, $W_r = 0$ i.e. total trust calculated by a neighbor is equal to trust calculated based on experience only.

4) Total Trust Calculated : Let the time duration for which the trust is calculated be from $(n-1)t_k$ to nt_k. Trust calculated for the given time duration is T_{tk} *L*et the trust value computed till the time $(n-1)t_k$ be called T_{prev}. T_{new} is the trust value computed from the previous knowledge and recent calculation. Considering, that the most recent event should have the greatest weightage, the total trust calculated by node x for node y can be given as:

$$T_{new} = \boldsymbol{\alpha} T_{prev} + (1 - \boldsymbol{\alpha}) T_{tk} \quad (7)$$

where $\boldsymbol{\alpha}$ is a decision parameter, which decides how much importance should be provided to past trust versus present time slot trust. The value of $\boldsymbol{\alpha}$ ranges from 0 to 1. The less the value of $\boldsymbol{\alpha}$, the more importance we give to the present events than the ones occurred in the past. As stated above, we give greater weightage to recent events and hence $\boldsymbol{\alpha} < 0.5$.

A node is usually deemed malicious if its trust is or falls below 0.3. Similarly, a certified authority could be implemented for nodes having trust greater than 0.75.

C. Trust in Ad Hoc Network using DSR

The trust levels for different nodes are computed and these values are added to the nodes in the ROUTE CACHE associated as weights to the link. Hence, by performing the shortest path algorithm, the sending node can find out the trustworthy node which may not be the best in terms of security nor the path may be the shortest one, but all the nodes carry along an associated level of trustworthiness with them.

We use the following method to find the best source route. Let A be a source node. On sending out a route request A expects to receive various route replies. Each of these replies comprise of a list of intermediate nodes that could serve as the route. Now the source node examines the trust of every node in a contesting route and checks whether it meets a certain threshold limit. We can use a route if trust through the route is higher than a threshold value.

Trust $_{route}$ = (trust of node 1) AND (trust of node 2) AND ….. (trust of last node in the route)

(T_a) AND $(T_b) = \min \{t_a, t_b\}$

However, to optimize the routing, all the routes that meet the trust threshold are once again tested to find which one of them is the shortest route to the destination. DSR normally achieves this using the Djikstra Algorithm. Introducing the trust threshold would secure the route and time for shortest path evaluation is also reduced because of reduced number of nodes, thus making our approach to source routing complete.

D. The Mobility Issue

An Ad hoc network is an ever-changing environment. The neighbors of a node may not remain constant in consecutive time slots. Thus precipitates a common scenario where a source route is broken due to loss of a neighbor in the route. The node which loses the neighbor has no other option than dropping the data packet and sending a route error message to the source. However, our trust management system deals judiciously with such a packet dropping node - considering its circumstances rather than any malicious nature.

In fact, the DTES resolves this problem efficiently. When the source receives a route error from node x, it sends an error acknowledgement (EA) out in the same route. This special packet informs all intermediate nodes before x in the broken source route to update their trust experiences for their neighbors. It may also contain distance weightage allocated to trust parameters measured from the source. A node y, preceding x in the route, had already received the route error. Now on receiving EA from the source, it updates trust for x based on its previous knowledge of x rather than punishing it for dropping the data packet.

E. Collusion Attacks

Collusion means a secret agreement between two or more parties for a fraudulent, illegal, or deceitful purpose. Such attacks are engendered by more than one malicious node and mainly consist of falsely recommending surrounding normal nodes (target nodes) through the recommendation parameter, either by over reputing a not very trusted node or under reputing a well reputed node. It may also include recommending a few target nodes very highly, so as to clog the network since all traffic will head to the target nodes. These cases are very easy to perform but very difficult to detect.

However, our trust management system takes care of such collusion attacks by *identifying anomalies in recommendations*. Clearly, when a malicious node defames or highly praises a target node, it is not following the network trend. In fact, its recommendation is contradictory to most of the true nodes. The trust management system performs the following algorithm to detect possible collusion attacks:

- On receiving recommendations at the end of the time slot, the trust calculator analyzes the received recommendations.

- Then, it tries to detect if one of the recommendations varies abruptly from most others.
- The abruptness is judged by a standard deviation function which decides the threshold of anomaly.
- If the threshold is crossed, the trust calculator node fixes a counter to the suspected recommender. From then onwards, this counter keeps track of the received recommendations for a fixed number of timeslots.
- Incase the anomalies continue to reappear in the suspect's recommendations; its trust is decreased by the trust calculator for suspected maliciousness.

IV. Simulation & Results

The simulations were designed keeping two factors in mind – Weightage and Thresholds, which appear throughout the paper and are very important ideas. Threshold imbues some tolerance in the system. Weightage ensures providing accurate importance to the parameters. Most of the trust calculation tests and formula verifications were done using step codes in Borland C++ environment. Performance testing was simulated using the network simulator ns-2 [18]. We decide on some *actual trust tendencies* of the nodes which is revealed progressively by long term behavior of the node.

Then the simulation begins. By sometimes using random or tendency specific values, we determine the number of successful and unsuccessful interactions for every node. These in turn are utilized in developing the trust of the nodes using DTES and TMS. As events proceed, our aim is to show that with passing events in the network, trust calculated using DTES and maintained using TMS tend to the actual tendencies of the nodes.

Scenario:

- 15 nodes are present in the entire ad hoc network.
- 8 nodes out of these 15 (1 – 8) are in the cluster under analysis.
- 3 nodes are malicious (nodes 3, 5, 7)
- Packet dropping at the rate of 8 packets/sec.

Tendencies:

Actual trust tendencies chosen were:

[0.44 0.55 0.3 0.4 0.2 0.8 0.26 0..58]

Sent data:

Source node = 7 destination node = 5 number of data packets sent = 1024 b

Source node = 7 destination node = 1 number of data packets sent = 2048 b

Source node = 2 destination node = 6 number of data packets sent = 2048 b

Source node = 4 destination node = 7 number of data packets sent = 2600 b

Source node = 6 destination node = 5 number of data packets sent = 5000 b

Number of Normal Test Packets Sent = 500 b

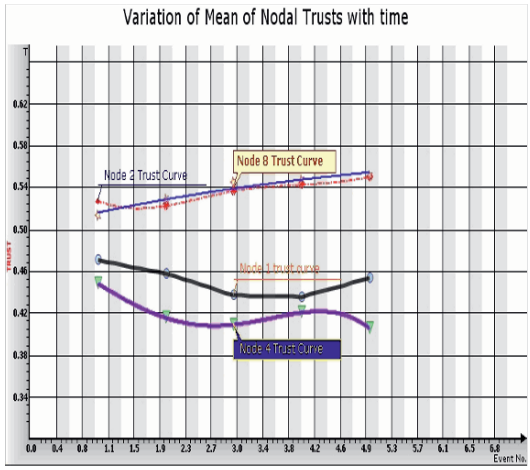


Figure 2. Variation of Mean of Nodal Trusts with time for nodes 1, 2, 4, 8.

The mean of the nodal trusts indicates the experimentally obtained tendency of the node. Nodes 1, 2, 4, 8 are not malicious. To enhance the view in Fig. 2, a y axis range of 0.3 – 0.66 is chosen. The figure clearly shows that extrapolation of the best fit lines obtained by plotting the mean trust values are tending towards the actual vector trusts for the nodes. For example, for node 2, the experimental value reaches 0.55 which is its actual value without need for extrapolation. On the other hand, node 8 best fit line if extrapolated would reach its actual value of 0.58.

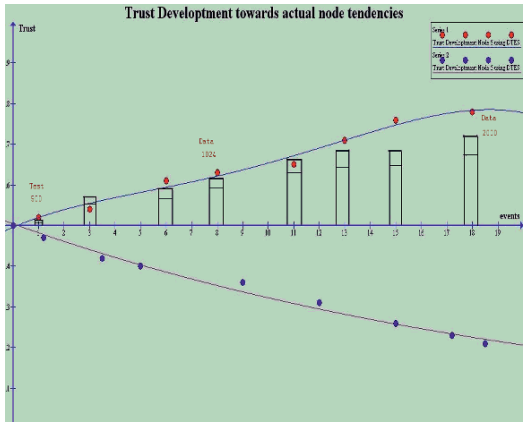


Figure 3. Malicious Node 5 and good node 6 trust development

Fig. 3 shows the trust development using DTES for nodes 5 and 6. According to the tendency, node 5 has a 0.2 trustworthiness and this is what the experimental results show thus proving our methodology. The bars associated with the good node 6 whose dynamic trust tends to 0.8 (Fig. 3), represent packets sent to it. The upper half of the bar represents the unsuccessful interactions while the lower half represents the successful interactions with node 6.

Our efforts are not to blacklist and remove malicious nodes. We aim to secure the DSR routing by ensuring that the source route contains trusted nodes. Malicious nodes can stay in the system, *but their presence will be nullified since they will never be chosen for routing purposes*. Hence, network performance is not diminished by deleting nodes. Instead, nodes are not selected in route unless they prove they are trustworthy. This is the vibrancy of our DTES.

V. Conclusion

Wireless ad hoc network are vulnerable to various types of attacks. Our aim is to develop such a trust management system that would invalidate the malicious nodes from an ad hoc community and provide secure DSR routing. We think that the technique is of practical significance since it can be implemented independently from routing software and does not require hefty modifications from the existing infrastructure. Our future work would be to secure specific attacks caused by malicious nodes and integrating energy constraints into trust computations.

References

[1] Dave B. Johnson and David A. Maltz, “The dynamic source routing protocol for mobile ad hoc networks”, Internet Draft, Mobile Ad Hoc Network (MANET) Working Group, IETF, October 1999.

[2] Indrajit Ray, Sudip Chakravorty, and Indrakshi Ray, Colorado State University, “VTrust: A trust management system based on vector model of trust”.

[3] Asad Amir Pirzada and Chris McDonald, School of Computer Science & Software Engineering, The University of Western Australia, “Establishing Trust in Pure Ad-hoc Networks”.

[4] L. Zhou and Z. J. Haas, “Securing Ad Hoc networks”, IEEE Networks Special Issue on Network Security, December, 1999.

[5] Fernandez-Gago, M. Carmen; Roman, Rodrigo; Lopez, Javier; “A Survey on the Applicability of Trust Management Systems for Wireless Sensor”, Security, Privacy and Trust in Pervasive and Ubiquitous Computing, 2007, SECPerU 2007, Third International Workshop.

[6] Venkatesan Balakrishnan and Vijay Varadharajan, “Designing Secure Wireless Mobile Ad hoc Networks”, Advanced Information Networking and Applications, 2005, 19th International Conference.

[7] Catharina Candolin and Hannu H. Kari,. “A Security Architecture for Wireless adhoc networks”, Helsinki university of technology.

[8] Pietro Michiardi and Refik Molva, “CORE: A Collaborative Reputation mechanism to enforce node cooperation in mobile ad hoc networks”, Proc. IFIP-Communication and Multimedia Security availability in mobile adhoc WANs, Proc. MobiHoc, Conference, 2002.

[9] Sonja Buchegger and Jean-Yves Le Boudec, “Nodes Bearing grudges: towards routing security, fairness, and robustness in mobile ad hoc networks”, Proc. 10th Euromicro Workshop on Parallel, Distributed and Network-based Processing, January 2002.

[10] Rajiv Misra, and C.R.Mandal, “Performance Comparison of AODV/DSR On-Demand Routing Protocols for Ad Hoc Networks in Constrained Situation”.

[11] Ali Aydin, Ersin Uzun, and Mark Pariente, “A Reputation based trust management system for P2P networks”, 2004 IEEE International Symposium on Cluster Computing and the Grid.

[12] Stephen Weeks., “Understanding Trust Management Systems”, IEEE 2001.

[13] Edith C.H.Ngai and Michael R.Lyu, “Trust and Clustering based Authentication Services in Mobile Ad Hoc Networks”, Proceedings of the 24th International Conference on Distributed Computing Systems Workshops (ICDCSW’04).

[14] Ren, Li, Wan, Bao, Deng, and Kim, “Highly Reliable Trust Establishment Scheme in Ad hoc Networks”.

[15] Florina Almen´arez, Andr´es Mar´ın, Daniel D´ıaz, and Juan Sanch´ez, “Developing a Model for Trust Management in Pervasive Devices”, Fourth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOMW’06).

[16] Asad Amir Pirzada, Amitava Datta and Chris McDonald. School of Computer Science, The University of Western Australia, “Trusted Routing in Ad-hoc Networks using Pheromone Trails”, “Propagating Trust in Ad-hoc Networks for Reliable Routing”.

[17] Yacine Rebahi, Vicente .E Mujica-V, and Dorgham Sisalem, “A Reputation-Based Trust Mechanism for Ad hoc Networks”.

[18] UC Berkeley and USC ISI, “The network simulator ns-2”, Part of the VIN T project. Available from http://www.isi.edu/nsnam/ns,1998.

[19] http://www.ietf.org, RFC – 177.

[20] http://www.ietf.org, RFC – 2441.

Investigating the Effects of Encoder Schemes, WFQ & SAD on VoIP QoS

Ajay Shrestha, Khaled M. Elleithy, Syed S. Rizvi
Computer Science and Engineering Department
University of Bridgeport, Bridgeport, CT 06601, U.S.A
{shrestha, elleithy, srizvi}@bridgeport.edu

***Abstract*- Voice Encoder Schemes, Weighted Fair Queuing (WFQ) and Speech Activity Detection (SAD) techniques affect the overall Quality of Service (QoS) of Voice over Internet Protocol (VoIP) services. VoIP is one of the most discussed and rapidly emerging technologies in telecommunication. We are slowly witnessing a change in telephony from Public Switched Telephone Network (PSTN) to IP based VoIP Network. Despite the benefits being enormous, the switch to VoIP hasn't been swift, primarily due to various performance (delay, jitter, packet loss, echo etc.) and security issues plaguing the VoIP telephony network. To achieve minimal QoS for telephony, the voice (packets) must be delivered within 150 ms to 200 ms. This paper presents a performance model to quantifying the influence of VoIP which gives an in-depth understanding of how Voice Encoder Schemes, WFQ and SAD influence VoIP QoS from a theoretical and implementation point of views.**

***Keywords*- VoIP, voice encoder schemes, voice activity detection, speech activity detection**

I. Introduction

VoIP is an upcoming technology, which will and already has revolutionized the way we communicate through telephony. VoIP is a vast subject and trying to touch all aspects of it is beyond the scope of this paper. Rather, I have focused on particular section of it. The paper is laid out in two parts:

- Literature Research
- Implementation: Encoders Schemes/WFQ Techniques/Speech Activity Detection

II. Related Work

In VoIP Network, an IP device (PC or IP enabled Phone) can make calls through the Internet. Here the call/voice bypasses the traditional switched network PSTN and travels as broken-down fixed-size independent IP packets through the Internet. Unlike in the switched network of PSTN, here each packet finds its own route and are reassembled in the correct order at the destination. The commercial feasibility and benefits from VoIP has fuelled its tremendous growth since the mid 1990s [5] and just the residential VoIP services is estimated to generate $4.1billion in 2010 [6].

VoIP Basics

We know that VoIP stands for Voice over IP (Internet Protocol), and IP indeed, is the most important aspect of VoIP. In fact the voice rides over IP, and this is how VoIP comes into existence. IP belongs to the Internet Protocol suite TCP/IP, which is the de-facto communication protocol for the Internet. In reference to the OSI model, IP equates to the Network Layer. As we know the network layer of OSI reference model is responsible for addressing, address resolution, routing, creating and maintaining routing tables and packet formatting. VoIP uses packet switching to transfer voice over the Internet, just as data is travels over the Internet. There are several technical details that need to be understood, to fully know how voice is carried though the data packet-switched network. The following background will give a good start to understanding it.

Pulse Code Modulation (PCM)

PCM is the way in which our analog voice is converted to a digital format, in the PSTN (Public Switched Telephone Network). Below are the steps that take place in PCM.

- First, the analog waveforms are filtered to remove anything greater than 4000 Hz, to remove any crosstalk from the voice signal. 0-4kHz is considered to be voice band.
- Then the filtered signal is sampled at 8000 times per second. The amplitude of the signal at the time of sampling is a 8-bit code.
- Since we are sampling at 8000 times a second. We have in hand 64000 bps. This is exactly how much the PSTN telephone infrastructure uses: 64kbps [1].

III. Pulse Code Modulation (PCM) in VoIP

PCM is also used in VoIP, but the bit code created is of different length for VoIP due to different voice compression methods (Voice Encoder Schemes) that are used. E.g. in G.729 voice compression technique, samples are taken at 8kbps and at that rate, it creates 10ms voice samples. By calculation each such sample works out to 10Byte (80 bits). Cisco IOS groups together two such samples in one packet. Also a header is attached to every packet [1]. Below is the calculation of total bandwidth required for such operation.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 501–505.

Table I
G.729 FRAMES PER PACKET AND BANDWIDTH

PCM Coding	G. 711	G. 726	G. 728	G. 729	G.723.1
Rate (Streams)	64 Kbps	16, 24, 32, 40 K bps	16 K bps	8 K bps	5.3, 6.3 K bps
5 being best and worst quality	4.1	3.85	3.61	3.7 – 3.92	3.65 – 3.9
User/Type	PSTN, PBXs Networks	PBX Networks	Low Delay Networks	Efficient ADPCM	Multi media service

A. *Bandwidth Calculation for G.729 Encoded Packet*

The following are the standard that we follow in our proposed solutions.

- G.729 samples at 8000 times per second creating a 8kbps code stream.
- 8kbps = 1KBps = 1 Bp(ms) i.e. 1 byte per millisecond
  Therefore every G.729 10 ms voice sample results in = 10Byte.
- By default 2 such voice sample are put in a Packet, so its gives 20Byte
- Thus 20 Bytes/frame worked out to 8kpbs.
- Add 40byte header to the packet. By above formula 40byte would require 16kbps. Therefore the total bandwidth required 8kbps G.729 codec = 24kbps
- There is an initial 5ms look-ahead delay (1st frame), thus Latency = 25ms

Using similar logic, Table I shows various results with variation of the G.729 parameters. Thus it can be said that, the lesser the number of samples per frame, the lesser the compression/packetization delay becomes, but it adversely consumes more bandwidth.

B. *Voice Encoder Schemes (Voice Compression)*

In PSTN 64 kbps PCM is used. Voice compression (codecs) uses several methods to compress the code so that less bandwidth is taken up. Codecs exploit repetitive characteristics in the voice wave to generate a compressed version of the waveform. There are several voice compression techniques, e.g. ADPCM (Adaptive Differential Pulse Code Modulation), CELP (Code Excited Linear Prediction Compression) and MP-MLQ PCM (Multi-Purpose Multi-Level Quantization PCM) [2]. Each of these techniques has their use in specific areas and condition. ITU-T (United Nation's governing body for Telecom Networks and Services) has grouped them in a series of recommendations named G-Series. Table II gives the G Series Coding Standards with its PCM streams rates.

Table II
PCM CODING G SERIES

G.729 Samples per Frame or Packet	IP/RTP/UDP Header	Bandwidth Consumed	Compression or Packet Delay
Default (2 samples per frame)	40 Bytes	24 K bps	25 ms
Satellite (4 samples per frame)	40 Bytes	16 K bps	45 ms
Low Latency (1 sample per frame)	40 Bytes	40 K bps	15 ms

Thus, we can see that the lower the Voice Encoder Scheme's Rate (streams), the quality of voice degrades accordingly.

C. *Voice Activity Detection (VAD) / Speech Activity Detection (SAD)*

Voice Activity Detection is an important part of the VoIP network. In a conversation, only one party talk at any given time, but today's network is made of bi-directional 64000 bps channel. Thus more than 50 percent (when accounting for breaks in speech) of the bandwidth is wasted, as voice is being sampled continuously irrespective if someone is speaking or not. SAD if enabled, can detect the magnitude (in decibels - dB) of speech and will stop voice from being framed if it detects no speech activity. Generally, SAD waits for a hangover time of 200ms for which there is no speech amplitude (decibels) before it stops putting the speech frames in packets. One inherent problem with SAD is that it cannot differentiate between noise and voice. The benefit of SAD is obvious, that the wasted (not used when party not speaking) bandwidth is put to use for something else.

D. *Quality of Service (QoS)*

QoS is the probability of meeting a given traffic contract, e.g. bandwidth and latency required for specific application. QoS can be broken down into CoS (Class of Service) and ToS (Type of Service). ToS is a field in the IP header that occupies 3 bits, enabling eight different types of CoS, 0-7. CoS categories packets into groups 0 through 7, depending on their bandwidth and latency requirements [4].

E. *Bandwidth Usage & Delay*

Bandwidth has always been the major concern with telephony, be it PSTN or VoIP. The voice-encoding scheme (codec) used and the number of voice samples per packet determines how much bandwidth is required for the VoIP network. Table III gives the bandwidth usage and delay for G.711 and G.729 encoder schemes depending on the samples per frame used.

Table III
BANDWIDTH USAGE VS CODECS (AND SAMPLES/FRAME)

Codec & Sampling Rate	G.711 (64Kbps)	G.711 (64Kbps)	G.729 (8Kbps)	G.729 (8Kbps)	G.729 (8Kbps)
Samples per Frame	one 10ms	two 10ms	one 10ms	two 10ms	four 10ms
Bandwidth	112kbps	96kbps	40kbps	24kbps	16kbps
Latency/Delay	10ms	20ms	15ms	25ms	45ms

One noticeable factor in reducing bandwidth usage is the number of samples used per frame, which is inversely proportional to the bandwidth usage. But negatively, the more samples you put in a frame, the more latency becomes. It can also be said from the above table that the bandwidth usage is less when Codec with lower sampling rate (stream) is used. Another conclusion that can be drawn from Table II and III is that, the quality of Voice starts degrading when we move to using lower bit rate stream encoders (codecs), thus the right balance of quality of voice and bandwidth usage needs to be sought, when choosing the Voice Encoder Scheme (codec), by looking at one's particular needs.

F. *Queuing*

As packets approach an interface (Router) for processing, they get queued while the processing is taking place depending on the nature of queuing algorithm used, the are released from the queue. The most simple of the queuing concepts would FIFO (First In First Out), whereby the packet that reaches the interface first gets to go out first. Taking the concept of Queuing to the next level, the packets can be classified into different categories and accordingly sorted into different priority queues. Packets in the higher priority queues pass though the interface faster than packets in the lower priority queues. In general there are three types of queuing method used: FIFO, Priority Queuing and WFQ (Weighted Fair Queuing)[3]. This paper focuses on WFQ.

G. *WFQ (Weighted Fair Queuing)*

WFQ differentiates traffic into several queues to separate flows and assigns equal bandwidth to each flow. This mechanism doesn't let one application (e.g. HTTP) to take over all the bandwidth. WFQ benefits low-volume applications allowing them to transfer faster while high-volume gets proportional amount of bandwidth. A good analogy to understand WFQ would be TDM (time-division multiplexing), whereby bandwidth is equally shared (by time-slots) between several channels or signal-streams. WFQ has an additional dynamic capability to sense absent data streams and then allocates that un-used bandwidth for other flows. In WFQ, streams are prioritized depending on the amount of bandwidth the flow consumes. So, basically the bandwidth is shared fairly by all applications. WFQ analyzes the source/destination address, socket/port number, protocol type and QoS/ToS (Type of Service) to determine flow type to categorize them accordingly. The weighting part of WFQ is determined by the following: IP Precedence, RSVP, IP RTP Priority, IP RTP Reserve, FECN (Frame Relay forward explicit congestion notification) and BECN (backward explicit congestion notification) [1]. FECN and BECN bits signify congestions, so such traffic is transmitted less. Values are assigned to each of the above factors, and the bandwidth allocated according to those values.

IV. PROPOSED IMPLEMENTATION AND SIMULATION RESULTS

Now that the literature research has covered the VoIP topics relevant to my research work, lets go into the implementation part. I've used OPNET IT Guru, the most widely used Network Simulation Tool in the academic arena, to perform my thesis implementation. Even outside the Academia, many small and large corporations use it alike. Even, Department of Defense (DoD) uses it for advanced Network Simulations.

A. *Simulation Tool and Specs*

For the sake of simulation, the OPNET IT Guru Academic Edition (9.1.A, build in 1996) is used. As a system specs, we use Windows XP Home Edition, service pack 2 with typically a small network using a mesh and bus topologies.

B. *Effects of Encoder Schemes and Speech Activity Detection on Load and Throughput*

This is the 1st part of my VoIP implementation. It demonstrates the effects of various voice encoder schemes on the load and throughput. As I mentioned earlier in the report, Encoder Schemes significantly affects the total bandwidth used by the link. Complex codec algorithms are used to reduce the sampling rate streams, which in turn reduce the bandwidth utilized.

Here, say a caller 0 and caller 1 from an office makes call to the another remote office. Caller 0 uses G.711 encoding on outgoing/incoming voice signal. Caller 1 on the other hand uses G.729 encoding on outgoing/incoming voice signal. Fig.1 shows the traffic (bytes/sec) received for both G.711 vs. G.729.

From Fig. 1, we see that traffic received for the Calling Party, is higher when G.711 (64kbps bit stream) encoder is used than when G.729 (8kbps bit stream) encoder is used. Thus we can conclude that when a higher bit encoder scheme

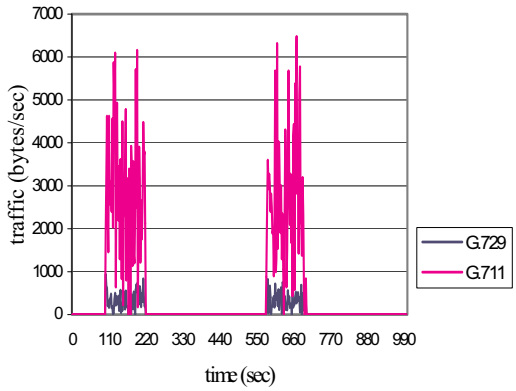


Fig. 1. Traffic Received: G.729 vs G.711

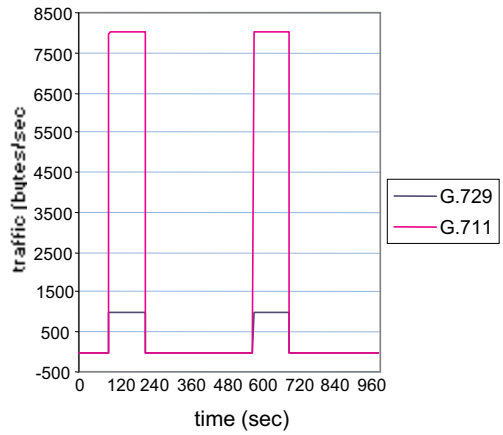


Fig. 2. Traffic Sent: G.729 vs G.711

is used, more voice traffic is generated and more bandwidth is required. Fig. 2 shows traffic sent for the different encoder schemes.

Similarly, from the above graph, we see that traffic sent by the Calling Party, is higher when G.711 (64kbps bit stream) encoder is used than when G.729 (8kbps bit stream) encoder is used. Thus like in the earlier case, we can conclude that when a higher bit encoder scheme is used, more voice traffic is generated and more bandwidth required, and thus adversely affects the VoIP QoS.

C. *Effects of Speech Activity Detection on Bandwidth*

Using the same setup as earlier implementation, the traffic generated for incoming and outgoing calls by the voice application is configured to be the same. Keeping that intact, now the traffic received is configured to use SAD. Fig.3 shows the traffic variation when SAD is enable and disabled.

In addition, Fig. 3 shows the Traffic Sent (without Speech Activity Detection) is higher than the Traffic Received (with Speech Activity Detection enabled) for the G.729 Application. Thus it can be concluded that enabling Speech Activity Detection lessens the traffic and frees up some bandwidth, and positively enhances VoIP QoS.

Fig. 3. G.729 Traffic Variation with & without SAD

D. *Delay Analysis With WFQ*

Now, that I've shown the effects of Voice Encoders and SAD on throughput and load, the next step my thesis implementation is to show the comparison of delay incurred, when WFQ is used with Voice on varying ToS (Type of Service) Applications. Here two nodes compete to send voice traffic through the same link between Router 1 & 2.

In the initial case, for both nodes the ToS (Type of Service) is set "best-effort" i.e. first-come first-serve basis. Fig. 4 displays the delay incurred by voice traffic from both these nodes.

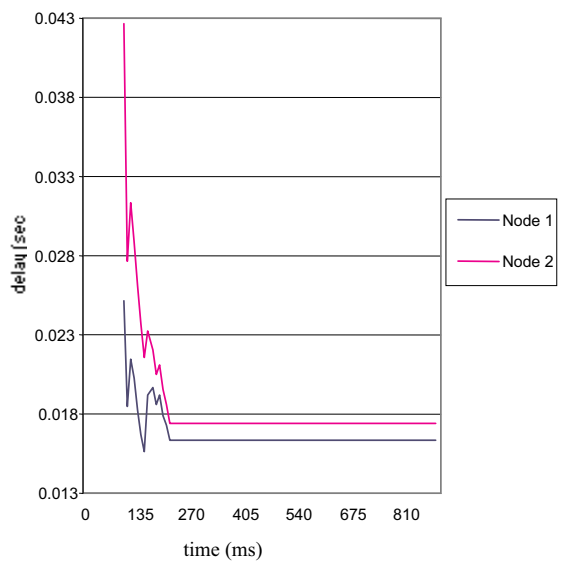
It should be noted in Fig. 4 that both nodes have almost same End-to-End delay when both are using the same ToS. The initial delay is slightly more for Node 1, primarily due to the variation in the traffic sent/received among the nodes.

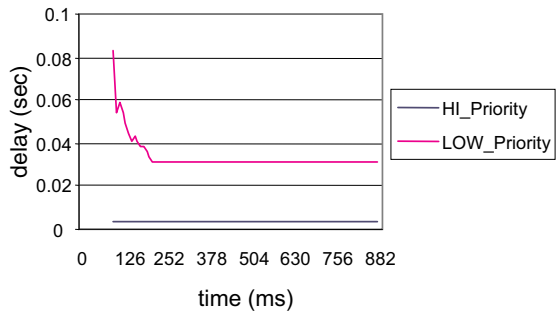
In the second scenario Node 1 is set to Hi_Priority i.e. the ToS is set to "Interactive Voice" and Node 2 is set to Low_Priority i.e. the ToS is set to 'excellent-effort'. Now the Router 1 and Router 2 will use the WFQ setting configured in the IP QoS attribute to prioritize the traffic from the two nodes.

Fig. 5 demonstrates the results of the proposed implementation. It should be noted in Fig. 5 that different ToS were used for the two traffic. 'Interactive Voice' has higher priority than the "Excellent-Effort" ToS value. Thus, HI_Priority_Traffic experience virtually no delay compared to the delay experience by LOW_Priority_Traffic. Thus we can clearly see that using WFQ significantly reduces the end-to-end delay and enhances VoIP QoS.

E. *Speech Activity Detection (SAD) and Bandwidth/Link capacity Utilization*

This implementation shows another aspect of how VoIP QoS is influenced by SAD. Speech Activity Detection greatly





helps make efficient use of available bandwidth. The 64kbps bi-directional voice channel's bandwidth is wasted more than 50% of the time due to break in conversations. So, SAD senses these breaks in conversations by keeping track of magnitude of speech (decibels) and uses the bandwidth for other traffic during the breaks in conversations.

In this implementation there are two calling nodes (say: Voice_src1 & Voice_src2) and two called nodes (say: Voice_dest1 & Voice_dest2). Voice_src1 and Voice_dest1 is one conversation pair and they use the G.711 voice encoder. Voice_src2 and Voice_dest2 is another conversation pair and they use G.729 voice encoder. The calling nodes are connected to router1 and called nodes are connected to router2, and router1 and router2 are in turn linked together. In the first simulation both conversation pairs use Speech Activity Detection (SAD) (also called silence suppression). In the second simulation, SAD is disabled. This will let us know the effects of SAD on bandwidth utilization. Bandwidth utilization will be shown as a total effect of both conversation pairs on the common link between router1 and router2. In addition, Fig. 6 shows how SAD affects the link bandwidth utilization. it can be seen in Fig. 6 that the point-to-point bandwidth utilization was optimally utilized and spanned a shorter period of time when SAD was enabled. So, it can be inferred that more calls can be made more efficiently utilizing the bandwidth, when SAD or silence suppression is used. Enabling Speech Activity Detection (SAD) detects notifies when either caller or called party is not talking (break in conversation for more that given amount of time), then the SAD will free up the bandwidth for other traffic. Thus SAD is an efficient way to utilize bandwidth and enhance VoIP QoS.

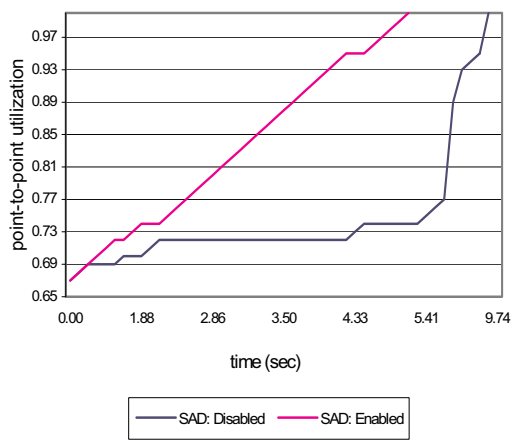


Fig. 6. SAD & Bandwidth Utilization

V. Conclusion

In this paper, we have explained how Voice Encoder Schemes, WFQ and Speech Activity Detection techniques affects the overall VoIP QoS in terms of Bandwidth and Delay incurred, from both an theoretical and implementation point of view. VoIP offers great benefits over the traditionally PSTN telephony, but it needs to achieve minimal QoS before it can completely replace the existing PSTN telephony. As more research and development work is being done on VoIP, it will only make is more viable for greater use and implementation in both residential and commercial telephony. Voice Encoder Schemes, WFQ and SAD, as described in this paper, are few of the major factors influencing the VoIP QoS, and this paper has in short but successfully show how it affects the VoIP QoS. The OPNET tool has been of tremendous assistance in visualizing the effects of the above factors on VoIP.

References

[1] Voice over IP Fundamentals, Tenth Printing April 2005, ISBN: 1-57870-168-6, Cisco Press, Jonathan Davidson, James Peters.

[2] Availability of Artificial Voice for Measuring Objective QoS of CELP CODECs and Acoustic Echo Cancellers, Paper by: Nobuhiko Kitawaki, Feng Wei, Takeshi Yamada, Futoshi Asano, University of Tsukuba, AIST, Japan, URL: http://wireless.feld.cvut.cz/mesaqin2002/full06.pdf

[3] QoS Queuing Techniques, Microsoft Corporation, Article ID : 233039, Last Review : October 30, 2006, Revision : 3.1, Resource URL: http://support.microsoft.com/kb/233039 as retrieved on Nov 18th, 2006 15:50:53 EST.

[4] Quality of Service Networking, Cisco Systems, Inc: URL: http://www.ciscosystems.com/univercd/cc/td/doc/cisintwk/ito_doc/qos.htm as retrieved on Nov 19th, 2006 13:22:03 EST.

[5] The Business of VoIP, Term Paper for 15.912, Technology Strategy, MIT Sloan School of Management, Jay Liu, Bassam Hajhamad, MBA Class of 2005, May 2005.

[6] World VoIP News, November 2006, Telecom Portal, Webiste, URL: http://www.voipproviderslist.com/world-voip-news/news/residential-voip-to-generate--41-bln-in-2010.html as retrieved on Nov 26th, 2006 17:33:30 EST.

[7] OPNET IT Guru Academic Edition, (9.1.A (Build 1996). URL: http://opnet.com/services/university/itguru_academic_edition.html Lab Manuals, as retrieved on Nov 28th, 2006 23:38:22 EST.

A Novel Approach for Creating Consistent Trust and Cooperation (CTC) among Mobile Nodes of Ad Hoc Network

Khurram S.Rajput, Khaled M. Elleithy, Syed S. Rizvi
Computer Science and Engineering Department
University of Bridgeport
Bridgeport, CT 06605
{krajput, srizvi, elleithy}@bridgeport.edu

Abstract- **This paper provides a critical analysis of the recent research wok and its impact on the overall performance of a mobile Ad hoc network. In this paper, we clarify some of the misconceptions in the understating of selfishness and miss-behavior of nodes. Moreover, we propose a mathematical model that based on the time division technique to minimize the node misbehavior by avoiding unnecessary elimination of bad nodes. Our proposed approach not only improves the resource sharing but also creates a consistent trust and cooperation (CTC) environment among the mobile nodes. We believe, that the proposed mathematical model not only points out the weaknesses of the recent research work but also approximates the optimal values of the critical parameters such as throughput, transmission over head, channel capacity etc. The simulation results demonstrate the success of the proposed approach that significantly minimizes the malicious nodes and consequently maximizes the overall throughput of the Ad Hoc network than the other well known schemes.**

I. Introduction

Misbehavior in mobile ad-hoc networks occurs for several reasons. Selfish nodes misbehave to save power or to improve their access to service relative to others [1]. Malicious intentions result in misbehavior as exemplified by denial of service attacks. Faulty nodes simply misbehave accidentally. Regardless of the motivation for misbehavior its impact on the mobile ad-hoc network proves to be detrimental, decreasing the performance and the fairness of the network, and in the extreme case, resulting in a non-functional network [2]. This paper addresses the question of how to make network functional for normal nodes when other nodes do not route and forward packets correctly. Specifically, in mobile ad-hoc networks, nodes do not rely on any routing infrastructure but relay on packets for each other. Thus communication in mobile ad-hoc networks functions properly only if the participating nodes cooperate in routing and forwarding. However, it may be advantageous for in nodes not to cooperate, such as a selfish node wants to preserve own resource to save power, memory, network-bandwidth, and local CPU time. Therefore nodes assume themselves that other nodes would forward the packet. This selfish or malicious intention of nodes can significantly degrade the performance of mobile ad-hoc-networks by denial of service.

In this paper, we focus on the design of a new time division based scheme that can avoid unnecessary elimination of malicious nodes while at the same time maximize the throughput of the system by increasing the recourse sharing among the mobile nodes. The existing methods/algorithms not only creating a performance bottleneck (i.e., by increasing the network congestion, transmission overhead etc.) but also diminishing the self-growing characteristic of a peer to peer network. These methods such as CONFIDANT [3] and CORE [4] force the participating nodes to adopt the same behavior as the other selfish nodes that have already been removed from the network due to the lack of resources. We believe that we should not propose any algorithm/method that becomes the reason for reducing the network resources and consequently force the existing participating nodes to behave exactly in the same way as other removed nodes. Instead, we strongly believe that we should come up with something that not only improves the resources and resource sharing but also creates a consistent trust and cooperation (CTC) environment among the mobile nodes.

The rest of this paper is organized as follows: Section II describes the research that has already been done in this area. The proposed analytical and mathematical models for CTC are presented in Section III. The simulation results are provided in section IV. Finally, section V concludes the paper.

II. Related Work

The terms *reputation* and *trust* are being used for various concepts in the literature, also synonymously [5, 6]. We define the term *reputation* here to mean the performance of a principal in participating in the base protocol as seen by others. The key thing in reputation system is watchdog and pathrater which have been proposed by Marti, Giuli, Lai and Baker [7]. They observed increased throughput in mobile ad-hoc networks by complementing DSR with a *watchdog* for detection of denied packet forwarding and a *path rater* for

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 506–511.

trust management and routing policy rating every path used, which enable nodes to avoid malicious nodes in their routes as a reaction. Their approach does not punish malicious nodes that do not cooperate, but rather relieves them of the burden of forwarding for others, whereas their messages are forwarded without complaint. This way, the malicious nodes are rewarded and reinforced in their behavior. They used a watchdog that identifies misbehaving nodes and a pathrater that helps routing protocols avoid these nodes. When used together in a network with moderate mobility, the two techniques increase throughput by 17% in the presence of 40% misbehaving nodes, while increasing the percentage of overhead transmissions from the standard routing protocol's 9% to 17%. During extreme mobility, watchdog and pathrater can increase network throughput by 27%, while increasing the overhead transmissions from the standard routing protocol's 12% to 24%.

CORE, a collaborative reputation mechanism proposed by Michiardi and Molva [4], also has a *watchdog* component; however it is complemented by a reputation mechanism that differentiates between subjective reputation (observations), indirect reputation (positive reports by others), and functional reputation (task-specific behavior), which are weighted for a combined reputation value that is used to make decisions about cooperation or gradual isolation of a node.

III. The Proposed Analytical And Mathematical Model For Creating Consistent Trust And Cooperation (CTC)

The creation of mathematical model can be viewed as a formalization of the proposed hypothesis. Based on the proposed mathematical model, we perform the numerical and simulation analysis for variety of scenarios in two parts. First, we use the mathematical model to run different scenarios in order to determine the performance of Ad-hoc networks by analyzing different critical network parameters such as throughput, transmission overhead and the utilization. Secondly, we use the same set of parameters as a performance measure.

A. The Proposed Analytical Model

We model the Ad-hoc network in much the same way as other researcher does except this paper introduces the new concept of time division. The idea of time division can simply be envisioned by considering a particular node of a network that has a potential to misbehave in the absence of the sufficient resources require to forward the packets of the neighboring nodes. This implies that if one can ensure that the network has enough resources that can be shared equally among the network nodes, then it can be assumed that the possibility of node misbehavior degrades significantly. Thus this reduction in the node misbehavior can be achieved through the time division technique that divides the time asymmetrically into the following two times: transmission-time required for *node-packets* and transmission-time required for *neighbor-packets*. The asymmetric division enables a node to effectively adjust the time required to transmit its own packets and/or the neighbor's packets. The reason for using the asymmetric division of the available time is to allow a node to effectively utilize the time by dividing it with respect to its current status (i.e., the available recourses) and consequently utilizing the bandwidth in an efficient manner.

The efficient utilization of the bandwidth satisfies the requirement of the fairness which is one of the key factors that forces a node to unfair with its neighbor. This indirectly points that we reduce the chances of misbehave since the node now has a total authority on the available resources. It should also be noted that we adopt an asymmetric approach to work with the time division method for this research which opposed to the conventional division of time (i.e., the symmetric or equal division employed by many different techniques). In other words, the proposed method optimizes the performance by effectively reducing the chances of node misbehave at the expense of comparatively complex logic.

B. The Proposed Mathematical Model

Before going to develop the actual mathematical model based on the above analytical model, it is worth mentioning some of our key assumptions. These assumptions help understanding the complex relationship between a large numbers of parameters. For the proposed mathematical model, we assume that a system has K nodes where each individual node k not only works as a normal mobile station but also works as a packet forwarding device for the other nodes. In addition, we assume that any kind of topology can be implemented among the mobile nodes to construct the Ad-hoc network. For the ease of simplicity, we perform the numerical analysis for a single node k. This can be further extended for the whole network by computing the collective behavior of the Ad-hoc network.

The primary principal of Ad-hoc network is that it allows each node of the network to fully participate in the construction of the network. The word *fully participation* leads us to the fact that a node not only transmits its own packets to the other neighboring nodes but also provides its services to other nodes as a forwarding device. For the proposed method, we assume that a node can decide to transmit its own packets with a certain probability while at the same time it can also deny the transmission of the other neighboring packets with a difference of a certain probabilities. In simple words, we can develop a relationship between these two probabilities as follows: *a node can transmit the self generated packet(s) with a probability of p where as it can transmit its neighbor packet(s) with the probability of q*.

Suppose, p is the probability for which a node forwards personal packets where as $p\ (1-p)$ is the probability for which a node transmit packets received from one ore more neighbors. In addition, we assume that k is total number of packets that can be transmitted by a certain node of the Ad-hoc network. The total numbers of packets include both the self generated packets and the packets receive from one or more nodes. Taking this into account, we can say that if the probability of transmission of a single packet is $(1-p)^x$ where x represents a single packet, then the probability to transmission

k packets would be $(1\text{-}p)^k$ where k represents the total number of packets that a node can transmit. This leads us to the following mathematical fact:

$$(1-p)^k \tag{1}$$

Equation (1) can simply be formalized for k number of packets as follows:

$$p(1-p)^k \tag{2}$$

As mentioned earlier, the proposed method is exclusively dependent on the time division methodology where a node can divide the time asymmetrically to represent the time it needs to transmit self generated packets as well as the time it takes to transit the packets arriving from one or more nodes. To make our proposed approach more realistic, we assume that if the packet that resides in a certain node is not delivered to its intended destination within the specified time, then that packet must be discarded by the node. The lost of the packet at the node level forces us to retransmit the packet. For the ease of understating, we assume that the time a node takes to transmit self generated packet can be represented as t_{pp} where as the time it takes to forward the packets received from one or more neighbors is represented as t_{np}. It should be noted that the total available time per node is just the sum of the time a node takes to transmit self generated packet and time it takes to forward the packets received from one or more neighbors. This relationship can be mathematically expressed in the following equation:

$$t_i = t_{pp} + t_{np} \tag{3}$$

where i represents the index of node that can be expended from 1 to K (i.e., K represents the total nodes present in a Ad-hoc network)

The maximum throughput is defined as the asymptotic throughput when the load is very large. In packet switched network where the load and the throughput are equal, the maximum throughput may be defined as the load in bits per seconds. Thus this in turns lead us to a fact that the maximum throughput can not be defined in the presence of packet drops at the node level. As mentioned earlier, to make our model more realistic we consider the possibility of packet drops and consequently the packet retransmission at the node level. The throughput from the proposed algorithm for a certain node of the Ad hoc network can be computed as follows:

$$T_{put} = Total\ Packets\ Forwarded \Big/ Total\ Time \tag{4}$$

The denominator of (4) is derived from (3) where as the numerator of equation is determined by using (1) and (2). One can see that as we increase the left hand side of (2), it causes a decrease in the left hand side of (4). It should also be noted that as we increase the sum of (1) and (2), it significantly increases the left hand side of (4). To make these relationships simple, we can say that the increase in the sum of (1) and (2) causes an increase in the throughput where as an increase in the total time that is determined by (3) causes a decrease in the throughput per node. This is because the more we increase the time, the more bandwidth we need to reserve to satisfy the transmission requirements.

A significant increase in the bandwidth utilization (which is beyond the scope of the available bandwidth per node) represents degradation in the throughput that indicates an increase in the possibility of node misbehavior. Thus, this implies that the proposed algorithm is not only improving the performance but also providing a chance to choose the optimal values of critical parameters. Equation (4) can be further simplified in the following form:

$$T_{put} = \frac{Node's\ Packets + Neighbour's\ Packets}{Total\ Time} \tag{5}$$

To formalize the above discussion, we can combine probabilities of transmission from (1) and (2) with the total available time per node from (3) in (5). Thus this expresses the node throughput not only by means of total available time but also by means of the total number of packets a node can transmit. The final result can be expressed in the following equation:

$$T_{put} = (1-p)^k + (1-p)^k \Big/ t_i \tag{6}$$

It should be noted that (6) gives node throughput by considering the time t_i spends on a single packet (that is the time spend on one packet is the sum of the time spend on self generated packets and the neighbor packets). Solving (5) for k number of packets in terms of the total time required by a node can be expressed in the following equation:

$$t_i = \sum_{1}^{k} t_{pp(k)} + \sum_{1}^{k} t_{np(k)} \qquad 1 \le k < \infty \tag{7}$$

where k in (7) represents the number of packets that are bounded between 1 and the infinity. The first and the second quantity of the right hand side of (7) are indicating the time required transmitting the self generated packets and the time required to transmit the neighbor packets. The generic time equation can simply be stated as:

$$t = no\ of\ packet \Big/ data\ rate \tag{8}$$

Using (8), one can now compute the two major components of the proposed time division algorithm. It is essential in order to understand the concept of asymmetric division. One of the two asymmetric time division quantities can be quantified as follows:

$$t_{np} = P(1-P)^k \Big/ D_R \tag{9}$$

where D_R in (9) represents the data rate.

Recall one of our fundamental assumptions that a node transmits k number of packets in total time t_i. This assumption allows us to set up a lower and upper bound on the number of packets that a node can transmit. Therefore, the limit for k should exist somewhere zero to infinity. One of the main reasons for recalling this assumption is make a more

generalized form of (9). Taking these two factors into account, one can generalize (9) as follows:

$$t_{np} = \sum_{k\geq 1}^{k\leq\infty} \frac{P(1-P)^k}{D_R} \qquad \text{where } 1 \leq K \leq \infty \tag{10}$$

The numerator of (10) is just a summation of total packets forwarded by a node with respect to the probabilities set up at static time. If t_{pp} is the total time taken by a node to forward its own k number ofpackets, then equation for t_{pp} can be rewritten as.

$$t_{pp} = \sum_{1}^{k}\left\{\frac{(1-P)^k}{D_R}\right\} \qquad \text{where } 1 \leq K \leq \infty \tag{11}$$

Equation (11) is the summation of probabilities of one packet to k number of packets per node in the presence of a certain data rate. By substituting the value of total time t_i from (3) into (6), we get

$$T_{put} = \left\{(1-p)^k + p(1-p)^k\right\} \Big/ \left\{t_{pp} + t_{np}\right\} \tag{12}$$

In order to generalize (12), we need to substitute the values of t_{pp} and t_{np} from (10) and (11), respectively, into (12), we get:

$$T_{put} = \sum_{1}^{k}\left[\frac{D_R\left\{(1-p)^k + p(1-p)^k\right\}}{(1-p)^k + p(1-p)^k}\right] \tag{13}$$

The first two quantities in denominator of (13) represent the summation of the time a node takes to transmit the personal packet and the neighbor's packets. It should be noted that (13) is generalized in a sense that it accommodates k number of packets that a node can deal at a certain point of time. To make it simple, we can rewrite equation as follows:

$$T(put\ of\ node) = \sum_{1}^{k}\left[\frac{D_R \times (1-p)^k + p(1-p)^k \times D_R}{\left\{(1-p)^k + p(1-p)^k\right\}}\right] \tag{14}$$

Equation (14) is the total throughput of a node for k number of packets that a node can transmit. Let us assume that Np is the power of node and K is the number of packet that a node can transmit. Taking these assumptions into account, one can derive a generic expression for utilization as follows:

$$U = N_{pout} \big/ N_{pin} \tag{15}$$

We call (15) as a generic mathematical expression of utilization, since both the numerator and the denominator are unknown and need to be determined to find out a more specific expression. Therefore, this new concept of power division leads us to the following mathematical expression for node-utilization with respect to the node's personal packets.

$$N_{ppout} = \sum_{1}^{K}\left\{K_{Pout}/t_{pp}\right\} \tag{16}$$

It should be noted that (16) is a more specific form of (15) since it only account for the personal packets. Thus the opposite hypothesis leads us to the following mathematical expression for the node utilization with respect to the personal packets:

$$N_{pnout} = \sum_{K\geq 1}^{K<\infty}\left\{K_{nout}/t_{np}\right\} \tag{17}$$

Contrary to (17), there should be an equivalent possibility of node inputs that can easily be computed as follows:

$$N_{pnin} = \sum_{1}^{K}\left(K_{nin}/t_{np}\right) \tag{18}$$

It should be noted that (18) can be useful to compute the output of the nodes in terms of the inputs of the node. In other words N_{Pout} is the sum of work on outgoing personal and neighbor packets that lead us to derive the simple mathematical relationship:

$$N_{pout} = N_{pp(out)} + N_{pnout} \tag{19}$$

In order to show that (19) is a valid true mathematical relationship between the input and output lines of a node, one needs to give another relationship as follows:

$$N_{pin} = N_{pnin} \tag{20}$$

This should now be clear that one of the reasons for deriving the above two relationship is to derive a more general expression from (16) and (17). Therefore, by substituting (16) and (17) into (19), we get the following equation:

$$N_{ppout} = \sum_{k\geq 1}^{k<\infty}\left[\left\{\frac{K_{ppout}}{t_{pp}}\right\} + \left\{\frac{K_{nout}}{t_{np}}\right\}\right] \tag{21}$$

Similarly, we can derive another expression using (20) which opposed to (21) as follows:

$$N_{Pin} = \sum_{1}^{K}\left(\frac{K_{nin}}{t_{np}}\right) \tag{22}$$

The last two equations (i.e., (21) and (22)) can now be used to derive the final expression for utilization as follows:

$$U = \sum_{1}^{k}\left[\left\{\frac{K_{pout}}{t_{pp}}\right\} + \left\{\frac{K_{nout}}{t_{np}}\right\} \Big/ \left\{\frac{K_{nin}}{t_{np}}\right\}\right] \tag{23}$$

All lines that are used for transferring the data or packets are also used for receiving the data or packets from neighbor nodes. This implies that the utilization per channel or line can be computed using (23). If we denote this line-utilization as (24), we can extend it to generalized (23).

$$U_R = \sum_{k\geq 1}^{k<\infty} \frac{\left(K_{pout}/K_{nout}\right)t_{np}}{K_{nin}} \quad (24)$$

If we assume that *n* numbers of routes are attached through the targeted node, then the utilization of the targeted node on all routes can simply be computed by summing the utilization of each node per channel. This can lead us to the following equation:

$$U_t = \sum_{n\geq 1}^{n<\infty} U_{R_n} \quad where\ 1 \leq n < \infty \quad (25)$$

This can also be interpreted as follows:

$$U_t = U_{R1} + U_{R2} + U_{R3} + \ldots\ldots + U_{Rn} \quad (26)$$

Therefore, the total utilization of system can be derived from (23) and (25) as follows:

$$U_t = \sum_{n\geq 1}^{n<\infty} \sum_{k\geq 1}^{k<\infty} \frac{\left(K_{pout}/t_{pp}\right) + \left(K_{nout}/t_{np}\right)}{K_{nin}/t_{np}} \quad (27)$$

We perform some simplification in (27) that results the following equation:

$$U_t = \sum_{n\geq 1}^{n<\infty} \sum_{k\geq 1}^{k<\infty} \frac{1}{K_{nin}} \left[K_{pout}\left(t_{np}/t_{pp}\right) + K_{nout} \right] \quad (28)$$

The above equation can be used to compute the total utilization of a certain node for all packets that it can forward and/or receive from one of its neighbor though all possible channels.

IV. The Experimental Verification And The Performance Analysis Of The CTC

We have shown that the system throughput can be measured in term of packets that neighboring node is generated as well as the self generated packets. To make the proposed methodology up to the standard, we derive the formula for computing the packet drop per node using (5). As mentioned earlier, we determine the behavior of the malicious node in terms of the number of packets that should have transmitted to the intended destination. For taking this into account, one can say that the effective throughput of a node is entirely dependence on how efficiently the node is forwarding the neighbor packets and thus creating a consistent trust environment among the nodes.

A. Case I

For case-1, we assume that the self generated packets per node are constant. We assume that one of the neighboring nodes of the target node sends packets at a certain rate that will increase linearly over the total simulation time. This assumption helps understanding the true performance of the proposed CTC algorithm. Fig. 1 shows the simulation results of packet-drops per node with respect to the number of packets generated by one of the neighboring nodes. It should be noted that as we increase the self generated packets, the number of packet-drops per node is increased. In addition, it can be seen in Fig. 1 that for a small value of neighbor packet generation (typically 500), both CTC and DSR are overlapping each other. However a slight increase in the neighbor packet generation causes a performance difference between these two approaches.

Figure 1: Neighbor packet generation vs. packet drop

B. Case II

CASE-II is different from CASE-I in such a way that both inputs of a node-forwarding system become a linear function of the node-time. The simulation result of this case satisfies the proposed mathematical model discussed in Section III in a way that the overall packet drop performance of both

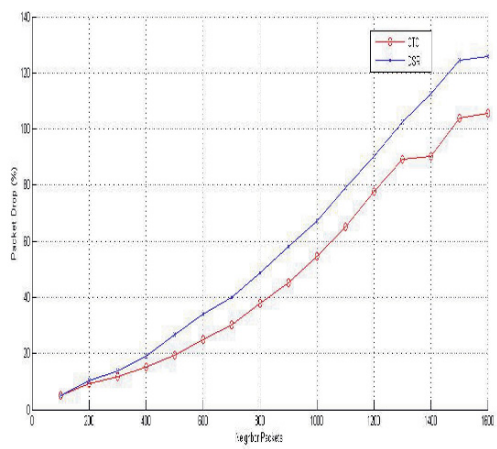


Figure 2: Neighbor packet generation vs. packet drop

investigated algorithms decreases. It can be seen that the packet drop is more rapid in Fig. 2 with respect to the neighbor-generated packets. In harmony with our expectations, as the number of neighbor-generated packets increased, the packet-drop performance of the proposed algorithm degraded. However, the performance degradation of the proposed algorithm was small compared to the performance degradation of the DSR algorithms.

C. *Case III*

The parameters-assumption for CASE-III is different from the previous cases in such a way that now one input (that is the neighbor-generated packets) of a node-forwarding system becomes a linear increasing function of the node total time where as the input (that is the neighbor-generated packets) becomes a linear decreasing function of the node total time. The expected output of this simulation was exactly the same as we were expecting based on our proposed mathematical model. That is the values of packet-drop for both CTC and DSR decreases as compared to the other two cases we discussed above.

D. *Case IV*

For this case, we assume that the neighbor-generated packet is a constant function of time. On the other hand, we consider self-generated packets as a linear increasing function of the total node time. It should be noted that the term linear increase or decrease implies a constant uniform change in the system parameter with respect to time. This case can also be considered as a reciprocal of CASE-I from its fundamental assumptions point of view. Thus we should also expect a reciprocal output for this simulation.

V. Conclusion

This paper proposed both analytical and mathematical model that can be used to effectively reduce the number of malicious nodes and packet drops. Our simulation results demonstrated that the proposed mathematical model not only points out the weaknesses of the recent research work but also approximates the optimal values of the critical parameters. Simulation results presented in this paper show that how the performance of mobile Ad hoc networks degrades significantly when the nodes eliminations are frequent. The simulation results of this paper are completely based on the proposed mathematical model for both lightly and heavily loaded networks. These results addressed many critical system parameters such as packet drop and packet loss versus malicious nodes, neighbor packet generation and drop ratio, and throughput per node per system.

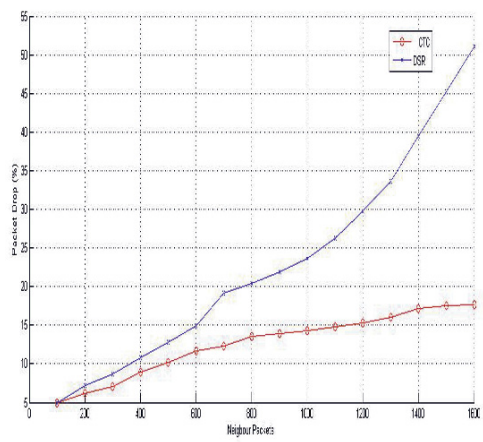


Figure 3: Neighbor packet generation vs. packet drop

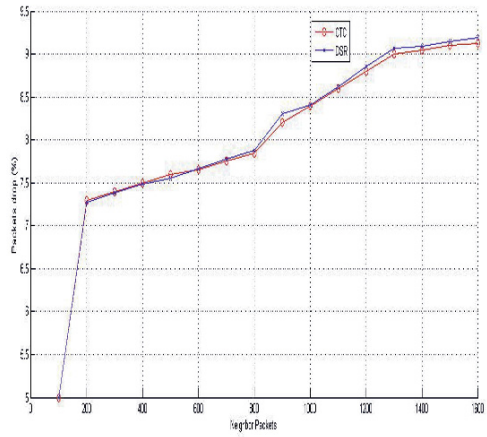


Figure 4: Neighbor packet generation vs. packet drop

References

[1] Zhang and W. Lee, "Intrusion Detection in wireless Ad-hoc networks," *in Proceedings of MOBICOM 2000*, pp. 275–283, 2000.

[2] Y. Huang, W. Fan, W. Lee, and P. Yu, "Cross-Feature analysis for detecting Ad-hoc routing Anomalies," *in Proceedings of the 23rd International Conference on Distributed Computing Systems (ICDCS 2003), Providence, RI*, pp. 478–487, May 2003.

[3] S. Marti, T. Giuli, K. Lai, and M. Baker, "Mitigating routing Misbehavior in mobile Ad hoc networks," *in Proceedings of MOBICOM 2000,* pp. 255–265, 2000.

[4] P. Michiardi and R. Molva, "CORE: A Collaborative Reputation Mechanism to enforce node Cooperation in mobile Ad hoc networks," *Sixth IFIP conference on security communications, and multimedia (CMS 2002)*, Portoroz, Slovenia, 2002.

[5] T. Moreton and A. Twigg, "Enforcing collaboration in P2P routing services," 2003.

[6] S. Bansal and M. Baker, "Observation-based Cooperation Enforcement in Ad hoc networks," Technical Report, 2003.

[7] S. Marti, T. Giuli, K. Lai, and M. Baker, "Mitigating routing Misbehavior in mobile Ad hoc networks," *in Proceedings of MOBICOM 2000*, pp. 255–265, 2000.

Bandwidth Problem in High Performance Packet Switching Network

Syed S. Rizvi[1], Aasia Riasat[2], Muhammad S. Rashid[3], and Khaled M. Elleithy[4]
Computer Science and Engineering Department, University of Bridgeport[1, 3, 4], Bridgeport CT, 06601
Department of Computer Science, Institute of Business Management[2], Karachi, Pakistan
{srizvi[1], muhammsi[3], elleithy[4]}@bridgeport.edu, aasia.riasat@cbm.edu.pk[2]

Abstract— **High performance packet switching networks are being deployed to provide sufficient data bandwidth for end users 3G services such as video streaming and broadband like data services. The use of high performance networks is, therefore, essential to the success of any 3G service. However, in practice, the deployment of High performance packet switching networks is hindered due to the improper congestion control which consequently results longer delays. In this paper, we propose a new method that can effectively improve the congestion control in high performance packet switching networks. Our numerical and simulation results demonstrate that the proposed method can be implemented for both lightly and heavily loaded networks. Simulation results show that the transmission delays can also be reduced significantly that improves the over all performance of high performance packet switching networks**

Keywords— high performance networks, 3G services, congestion control, transmission delays

I. Introduction

As in Packet Switching Networks the message is decomposed into packets which share channels with other packets [3]. Packet high performance networking strives to achieve the highest possible latency, high utilization of the expensive resources, fair allocation of resources to competing users (QOS guarantees), or combinations of all these. Actually, Packet Switching network was originally designed to provide more efficient facility than circuit switching for bursty data traffic. Here the bandwidth seems utilized more efficiently. Main advantages of Packet Switching are flexibility, resource sharing, robustness, responsiveness.

Today's internet is built up as a giant packet switched network. Since packet switching offer better bandwidth sharing and is less costly to implement than circuit switching, it is widely used in data networks [1]. The internet transmits data packets efficiently but it provides no guarantee on the end to end transmission delay (up bound, jitter etc). Packet losses take place due to congestion [2, 3]. Moreover delay predictions are very important for the design of routing and flow control algorithm [4, 5].

A. Problem Statement

Congestion is a problem that occurs on shared networks when multiple users vie for access to the same resources (bandwidth, buffers and queues) [5]. In packet switched network, packets move in and out of the buffers and queues of switching devices as they traverse the network. In fact, a packet switched network is often referred to as a "network of queues". A characteristic of packet switched networks is that packet may arrive in bursts from one or more source. Buffer help routes absorb bursts until they can catch up. If traffic is excessive, buffers is not a solution because too much buffer size can lead to excessive delay.

The current analysis of congestion control algorithms shows that as the combination of both delay and the bandwidth increases, the transmission control protocol becomes oscillatory that makes the overall system slightly instable. Recently, reference [11] shows that the increase in the product of capacity and delay results instability in a random early discard [7], random early marking [6], proportional integral controller [9], and virtual queue [8]. In addition, it has been shown that an active queue management scheme can maintain stability over very high-capacity or large-delay links. Also, it has also been shown that the adaptive virtual queue [10] also becomes prone to instability when the link capacity is large enough.

Congestion occurs when various sources compete for network resources, and these resources cannot handle the demand [3]. This may happen when logical channels request bandwidth that cannot be supported, or when the network admits more packets than the links can handle, or at any node due to buffer shortage. Here we select problem called congestion so that we can make packet switch network more efficient. Congestion problem leads to packet loss so if we can minimize its effect then we can improve the performance of high performance packet switched network.

Congestion in a network occurs because routers and switches have queues buffers that hold the packets before and

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 512–515.

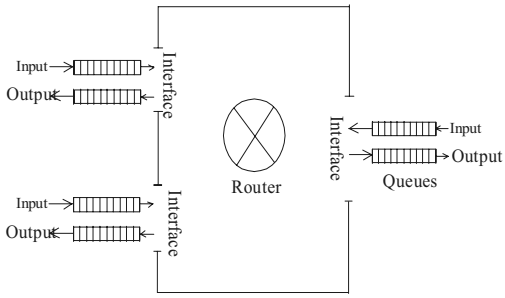


Fig. 1. Queues in the Router

after processing. A router for example, has an input queue and an output queue for each interface .When a packet arrives at the incoming interface .It undergoes three steps before departing.

- The packet positions at input queue while waiting to be checked.
- The processing module of the router removes the packet from queue and uses its routing table to make decision.

The packet is plant in the appropriate output queue and waits its turn to be sent.

II. Related Work

Much research has been done to lessen the congestion problem in the network. Among them two methods to control the congestion are described under.

A. Backpressure Mechanism

The technique of backpressure refers to a congestion control mechanism in which a congested node stops receiving data from the immediate upstream node [2, 7]. This may cause the upstream node to become congested, and they in turn, reject data from their upstream nodes and so on. Backpressure is a node to node congestion control that starts with a node and propagates, in the opposite direction of data flow to the source. The node 2 will have the congestion so it will inform node one to slow down. If congestion occurs in node 1 then it will inform source to slow down. This is how backpressure works.

B. Choke Packet Mechanism

A choke packet is a packet sent by a node to the source to inform it of congestion. In choke method the warning is from the router, which has encountered congestion to the source station directly [4, 11]. The intermediate nodes through which the packet has traveled are not warned. Instead the source are warned about the congestion occurs in the router.

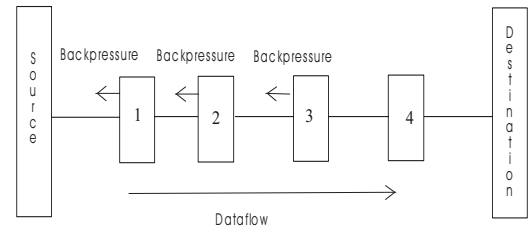


Fig. 2. Backpressure method

III. Proposed Solution For Traffic Management In ATM Networks

Last two solutions loss the data. We have proposed the elucidation without any loss of data. Here we made some assumptions, they are as follows:

- The sender and receiver know the size of the buffer the router have.
- The packet size is known by both.

The sender and destination knows the maximum time for processing the maximum size of buffer the router has. In Fig.5 the router informs the congestion to the source rather than node. Firstly the source sends the packet to the router. As the source knows the buffer size of the router and its processing time, In addition to the packet it also sends the control bit. The control bits are sent from the router to sender in certain interval. Here, we made the assumption that if the control bits from the router send that the buffer is 70% full then the sender will stop sending the packets. If the buffer size of router is 50% full then the sender will slow down by 25%.Similarly if router buffer is even less than the 50% then the sender will send the packet in full speed. After certain interval the control

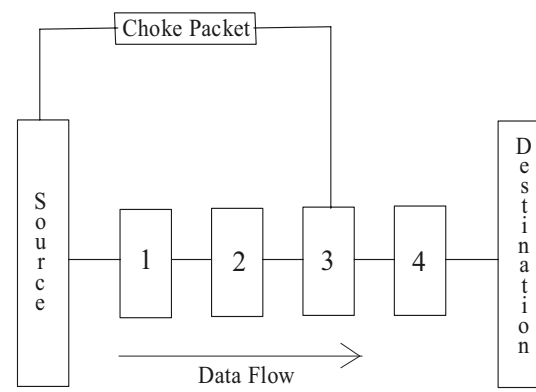


Fig. 3. Choke Packet Mechanism

bits are broadcasted to each sender. So by looking at the buffer usage percentage we can slow down the speed. The more it takes time to return the slower will be the packet delivery from the source. Here we take n number of senders.

Now, lets move to the congestion occurs in output side of the router as shown in Fig. 6. We know the destination knows time of processing of the packet as well as the time to get that packet which is sends by the source. Suppose the destination didn't get the packet in time then the destination will send the control bit. If control bit takes time to come back to the receiver. The receiver will know there is congestion occurring in the output buffer of the router. This implies that, if router gets control bit from the receiver the router is going to slow down the processing speed so that there will be time for the receiver to get the packet without loss.

IV. Proposed Mathematical Model

Before going to present the mathematical model, it is worth mentioning some of our assumptions:

- First, we consider both sender and router part.
- We assume that we may have *n* number of senders
- Also, we assume that the length of the link is L_n whereas the sender is represented by S_n

Taking this into account, we can present the following mathematical expression to formulate out proposed hypothesis as follows:

$$\text{Total no of senders} = \sum_{n=1}^{\infty} S_n \qquad \text{Property (1)}$$

In the same manner, we can compute the transmission time as follows:

$$\text{Transmission Time} = Ts = L / B \qquad \text{Property (2)}$$

where L is the length of the packet (typically measure in bits) and B is the bandwidth of the link (typically measure in bits/sec). Combining property (1) with (2), results the following mathematical expression: In addition to that, we also

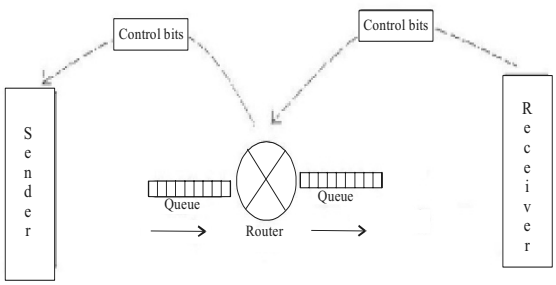


Fig. 4. Congestion control technique

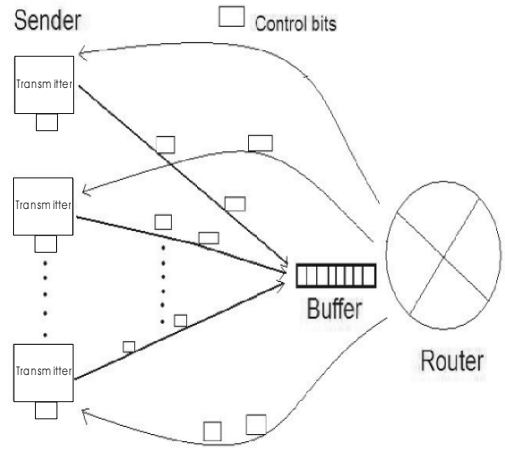


Fig. 5. Sender and Router Communication packet

assume that the propagation time remains constant for both transmission of messages and represented by T_p. With this assumption, the total time taken by *n* number of senders to send first bit will be approximated as follows:

$$.T_s(total) = \sum_{n=1}^{\infty} S_n \times T_n \qquad \text{Property (3)}$$

$$\text{Router} = \sum_{n=1}^{\infty} \left(T_s\ (total)\ +\ T_P\right) S_n \qquad \text{Property (4)}$$

According to one of our assumptions, there are n no of receivers whereas the length of the link between the router and the receiver is assumed to be Lr. It should also be noted that the same link for the receiver side is used and represented by DSI. Taking these factors into account and combine them with the four properties presented above, we can present the following mathematical expression for the total number of receivers.

$$\text{Total no of receiver} = \sum_{n=1}^{\infty} Rn \qquad \text{Property (5)}$$

Also, we assume that the *Processing-Time* (P_r) that router takes is Rt. In addition, this time is assumed to be known by both sender and the receiver. This implies that the time required to transmit a single bit is a sum of processing time and the router processing time. This can be mathematically expressed as:

$$Trnamission - Time / bit = P_r + R_t \qquad \text{Property (6)}$$

This can be generalized for *n* number of receivers

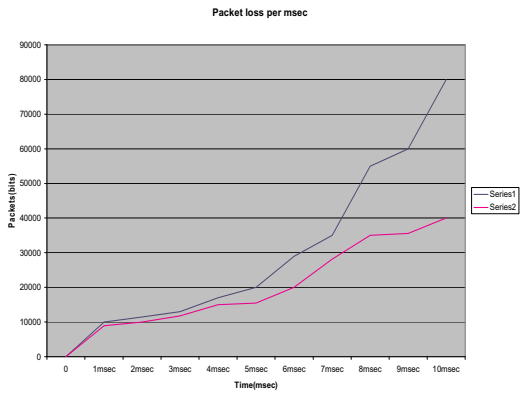


Fig.6. packet loss versus time (msec)

$$R_n = \sum_{n=1}^{\infty} (P_r + R_t) R_n \qquad \text{Property (7)}$$

$$where\ R_n \xrightarrow{approximate} n\#of\ \text{Receivers}$$

V. Simulation Results of the Proposed Scheme

Fig. 6 represents an approximation of packet loss with respect to time. As harmony with Fig. 6, an average packet loss of the presented model can be numerically approximated using above properties:

$$(Ts) = 1000/1544 = 0.647\,\text{sec}.$$

On the other hand, average packet loss of previous model is typically presented as: $Ploss(old) = 30045.45$ .This further leads us to the following performance analysis.

$$performance = \frac{Ploss(old) - Ploss(present)}{Ploss(old) * 100} \hat{=} 33.4\%$$

$$where\ \hat{=}\ \text{represents estimated value}$$

Here, we can take out the transmission time, we are using DS1 link in between the router and the sender. So the bandwidth is 1.544 Mb/sec and frame or packet size to be 1000kbits.

VI. Conclusion

There are lots of things to be considered for the control of congestion problem. Either you have to compromise with the time, cost or packet loss. In our model, we have compromised with the time and cost but put effort to minimize the packet loss. As we all know that packet switching network is based upon connectionless network. Due to this there is lots of packet loss. According to the previous model we found loss of data to be maximum. So we somewhat tried to diminish the packet loss by making some of the assumptions.

References

[1] L. Klein, S. Lam, "Packet Switching in a Multi-access Broadcast Channel: Performance Evaluation," *IEEE Transaction on Communications*, Vol. 23, Issue. 4, pp. 410- 423.

[2] S. Kalyanasundaram, J. Li, E.K.P. Chong and N.B. Shroff, Channel sharing scheme for packet-switched cellular networks, in: *IEEE INFOCOM'99* New York (March 1999) pp. 609-616.

[3] B. Vandalore, W. Feng, R. Jain, S. Fahmy: A Survey of Application Layer Techniques for Adaptive Streaming of Multimedia. Real-Time Imaging 7(3): 221-235 (2001)

[4] H. Jiang and S. Rappaport, CBWL: A new channel assignment and sharing method for cellular communication systems, *IEEE Transactions on Vehicular Technology*, Vol. 43, Issue. 2, pp. 313-322, 1994.

[5] R. Jain, K. Ramakrishna, D. Chiu, "Congestion Avoidance in Computer Networks with a Connectionless Network Layer," *Innovation in Internetworking*, pp. 140 – 156, 1998.

[6] S. Athuraliya, V. Li, S. Low, and Q. Yin, "Active queue management," *IEEE Network*, 2001.

[7] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *In IEEE/ACM Transactions on Networking*, Vol. 1, Issue. 4, pp. 397–413, Aug. 1993

[8] R. Gibbens and F. Kelly, "Distributed connection acceptance control for a connectionless network," *In Proc. of the 16th Intl. Telegraffic Congress,* June 1999

[9] C. Hollot, V. Misra, D. Towsley, and W. Gong, "On designing improved controllers for AQM routers supporting TCP flows," *In Proc. of IEEE INFOCOM*, Apr. 2001

[10] S. Kunniyur and R. Srikant, "Analysis and design of an adaptive virtual queue," *In Proc. of ACM SIGCOMM*, 2001.

[11] S. Low, F. Paganini, J. Wang, S. Adlakha, and J. Doyle, "Dynamics of TCP/AQM and a scalable control," *In Proc. Of IEEE INFOCOM,* June 2002.

An Efficient Scheme for Traffic Management in ATM Networks

Syed S. Rizvi[1], Aasia Riasat[2], Muhammad S. Rashid[3], and Khaled M. Elleithy[4]
Computer Science and Engineering Department, University of Bridgeport[1, 3, 4], Bridgeport CT, 06601
Department of Computer Science, Institute of Business Management[2], Karachi, Pakistan
{srizvi[1], muhammsi[3], elleithy[4]}@bridgeport.edu, aasia.riasat@cbm.edu.pk[2]

Abstract— **As ATM network is designed for broad band transmission that is high data rate (25 Mbps to 2.5 Gbps) and supports the transmission of every kind of data, congestion control and delay have been important issues for ATM networks. Data transmission is done in the form of cell (53 bytes) relay. Hence, cell sequence and the error control have to be carried out properly. ATM networks presents difficulties in effectively controlling congestion not found in other types of networks, including frame relay networks. In this paper, we present an efficient methodology for traffic management. The simulation results suggest that the proposed solution is effective for both slow and high data rate transmission.**

Keywords— ATM network, broad band transmission, congestion control, transmission delay

I. Introduction

International Telecommunication Union (ITU) has defined a restricted initial set of traffic and congestion control capacities aiming at simple mechanisms and network efficiency as follow. That sets the upper bound to the traffic, variability in the pattern of cell arrival and average rate of ATM connection.

1. Connection Admission Control.
2. Usage Parameter Control.
3. Priority Control.
4. Fast Resource Management.
5. Selective Cell Discarding.

Apart from that, ATM switch is important network device in ATM network for congestion and traffic control. That switches virtual circuit identifier (VCI) from left to right. It contains buffers and switching circuits to guide the connections.

A. Problem Statement

Traffic from user nodes can exceed the capacity of the network, which causes memory buffer of ATM switches to overflow and data loses. As per high data rate of transmission, cell storage and traffic management is required in ATM switches.

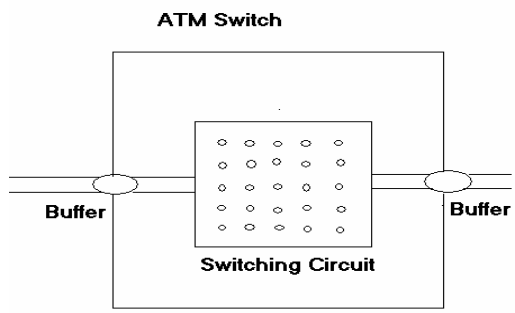


Fig. 1. Switching Architecture

The restrictions of ITU-T can be managed by upgrading the performance of ATM switches. ATM switches takes a time to process each cell's VCI through switching circuit and referring to the routing table instead of no time. That affects the continuity of the cell transmission.

ATM switches do outgoing buffering if more than one cells have same VCI, which leads to defer in transmission and affect the CBR (Constant bit rate). That results in retransmission or poor performance at receiver side especially in case of video-audio data.

II. Proposed Solution For An Efficient Traffic Management In ATM Networks

Small buffers of ATM switches can be replaced with large memory blocks.

A. Replacing Switching Circuit

ATM switching circuit can be replaced with processing unit, which has capability of processing cells, just like CPU. By processing unit cells are transmitted faster or no time and provides continuous flow of transmission. Routing table is allocated in memory of processing unit. That does congestion

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 516–518.

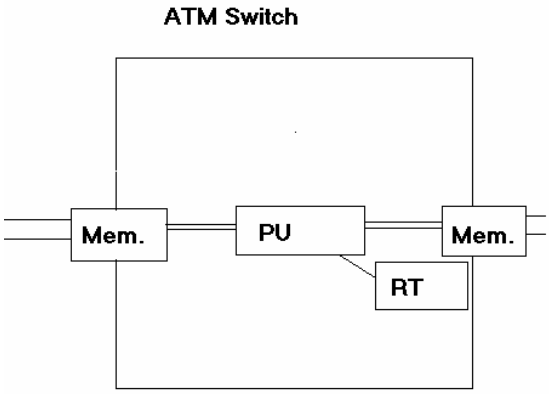


Fig. 2. A conventional ATM switch with replacement of switching circuit

control and speeds up the transmission of outgoing cells so that less buffering is required. Due to less buffering and no data lose, retransmission is not needed to perform.

B. Removing Output Buffers

Another important issue, to fasten the transmission is to prevent the output buffering. ATM switches buffer the out going cells, if they have same VCIs. Processing unit can process those cells at a time that have different VCIs. And the resulting (outgoing) VCIs of those cells will be different. Hence, outgoing cells need not to be buffered

III. Proposed Solution For Traffic Management In ATM Networks

A. Scalability

Moreover, if there is higher bandwidth, it is needed to minimize the buffering. For that, processing unit can be scaled through transmission in ATM switch. More than one processing units are implied in ATM switches. At a same time, more that one cell can be processed. Incoming flow is divided into the number of channels and assigned to the processing units having their own memory blocks.

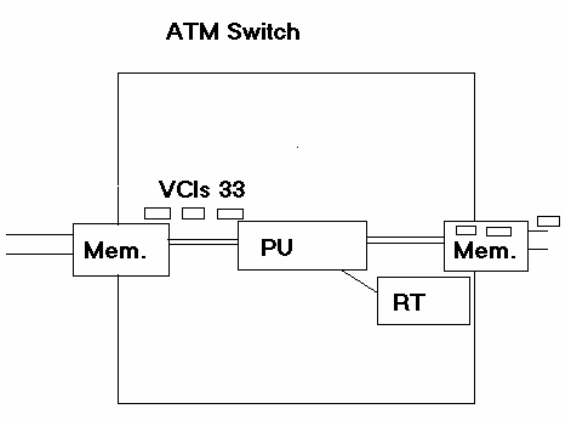


Fig. 3. A conventional ATM switch without output buffer

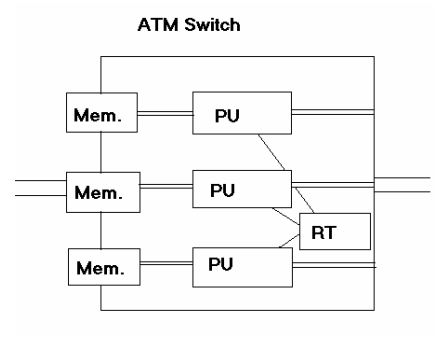


Fig. 4. An scalable ATM switch architecture

PU- processing unit.
RT- routing table.
Mem.- memory block.

IV. Proposed Offset Mechanism In Routing Table

Processing unit may take some execution time to access the outgoing VCI. Instead of searching through whole routing table, table can be divided into segments having offsets. Each cell can refer to addressing table to access the offset. And using that offset outgoing VCI can be defined. Thus, instead of searching nine rows, each cell has to refer to only three rows. That way it fasten the access of VCIs and transmission of cells.

A. Error Control Integration

Secondly, HEC algorithm can be implemented in ATM switch instead of in receiver. If data is corrupted in cells, that will be discarded by ATM switches. That results in improvement of quality of service in transmission.

B. Mathematical Model

CBR of network depends upon the constant duration of time between arrivals of two subsequent cells to receiver.

That, cell duration is defined as

$$\delta = 1/R \quad (1)$$

Where R= Data rate.

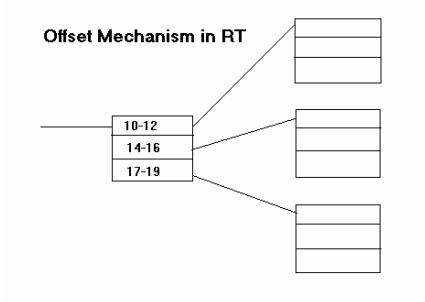


Fig. 5. A proposed approach for offset mechanism for ATM networks in RT

Hence, as data rate is higher, quality of service and CBR can be improved. And higher cell duration (δ) affects the quality of service badly.

CBR can be defined as,

$$CBR \rightarrow C\ R \quad (2)$$

Where R= Data rate, C= Channel capacity

Case- 1
Suppose if the velocity of cell across the medium approaches 300 km/s, then the following conclusion can be drawn:

Average holding time of ATM switch with switching circuit for one cell > 1 s = 1.4 seconds (or more than transmission medium). This leads us to the following expression:

That decrease velocity and data rate as, $R \rightarrow V$

$$\delta \rightarrow 1/R \quad (3)$$

In (3), we can conclude that the values for δ (cell duration) will be increased with respect to time. Thus, as ATM switch with processing unit has less average holding time for cells, it provides improved quality of service and CBR by formula (2).

Case- 2
Channel capacity of medium is defined as,

$$C \rightarrow SNR \text{ and } BER \rightarrow 1/SNR \quad \text{Property (4)}$$

Hence, as we decrease the bit error rate, signal to noise ratio and channel capacity can be improved. That results in better quality of service and CBR by formula (2).And as per the solution, ATM switch with processing unit reduces the bit error rate by HEC algorithm and by reducing possibility of retransmission.

V. Conclusion

In this paper, we have investigated different architectures for ATM switches to improve the overall performance of ATM networks. We have shown that an appropriate architecture for ATM switch can provide a strong congestion control which can consequently use to improve the traffic management in ATM networks. In addition, we described that how scalable performance can be achieved from the ATM networks if we deploy offset mechanism in ATM switches.

References

[1] Abry, P. & Veitch, D, "Wavelet analysis of long range dependent traffic," *IEEE Transactions on Information Theory*, Vol. 44, Issue. 1, 1995.

[2] Bjorkman, N., Latour-Henner, A., Hasson, U., Pers, O., & Miah, A., "Practical ATM resource dimensioning based on real-time traffic measurements and analysis," *GLOBECOM*, Vol. 1, pp. 399-403.

[3] Monlar, S. & Vidacs, A., "On modeling and shaping self-similar ATM traffic," TR. High speed Networks, Laboratory, Department of Telecommunications and Telematics, Technical University of Budapest.

[4] G. Stamoulis, M. Anagostu, and A. Georgantas, "Traffic source models for ATM networks," *A survey in Computer Communications*, Vol. 17, Issue. 6, pp. 428-438, 1994.

[5] Willinger, W., Taqqu, M.S., Sherman, R. & Wilson, D.V., "Self-similarity through high variability: Statistical analysis of Ethernet LAN traffic at the source level," *IEEE/ACM Transactions on Networking*, 2004.

Use of Self-Adaptive Methodology in Wireless Sensor Networks for Reducing the Energy Consumption

Syed S. Rizvi, Khushboo Patel, and Chaitali Patel,
Computer Science and Engineering Department
University of Bridgeport, Bridgeport, CT 06601
{srizvi, khushbop, chaitali}@bridgeport.edu

Abstract - **A sensor network is made up of numerous small independent sensor nodes with sensing, processing and communicating capabilities. The sensor nodes have limited battery and a minimal amount of on-board computing power. A self-adaptive methodology that utilizes the source and path redundancy techniques to efficiently reduce the required energy consumption is proposed. The proposed methodology presents a self-adaptive strategy to optimize the number of active sensor nodes and assign equal time slots to each sensor nodes for sensing and communication with the BS. The self-adaptive property enables the proposed methodology to perform a global search for optimizing the position of active sensor nodes with respect to the location of the base station (BS). The global search performed by the proposed methodology is carried out in a complete top down manner. The proposed self-adaptive methodology, therefore, not only reduces the energy consumption of wireless sensor nodes but also effectively maximizes the lifetime of active sensor nodes. Simulation results demonstrate that the proposed methodology significantly minimizes the energy consumption and consequently increases the life time of active sensor nodes.**

Keywords - **Base station, energy consumption, self adaptive methodology, sensor node.**

I. INTRODUCTION

A WSN is a network made of numerous small independent sensor nodes to monitor environment at different locations. The sensor nodes, typically the size of 35mm film canister, are self-contained units consisting of a battery, radio, sensors, and a minimal amount of on-board computing power. Nodes must have self configuration and adaptation mechanisms to support fault tolerance.

In the past few years, the rapid development in miniaturization, low power wireless communication, micro-sensor and small-scale energy supplies have made WSNs a new technological vision. WSNs show great potential for increasing the information available to people in a wide variety of consumer and industrial applications.

While a lot of research has been done on some important aspects of WSNs such as architecture and protocol design, energy conservation, and localization, supporting Quality of Service (QoS) in WSNs is still a largely unexplored research field. This is mainly because WSNs are very different from traditional networks. Communicating one bit over a wireless medium at short ranges consumes more energy than processing that bit. With the current technology, the energy consumption for communication is several magnitudes higher than the energy required for computation; and wireless communication is foreseen to continue to dominate energy consumption in the near future. There are two possible ways to decrease the energy used for communication in a sensor network: minimize the amount of the transmitted data, or by decreasing the communication range. The transmission energy is proportional to d^{α}, where d is the transmission distance, while α is the attenuation exponent. Thus, minimizing the amount and the range of communication as much as possible can significantly prolong the life of a sensor network.

In this paper, we propose a *self-adaptive methodology* which takes into consideration the above proposed possible ways to increase the lifetime of the sensor networks. In [5], the concept of moving the BS to its optimum position is introduced. In addition to position optimization, we propose a novel *self-adaptive methodology* that optimizes the number of active sensor nodes. Simulations results demonstrate that the proposed *methodology* significantly reduces the energy consumption while at the same time provide a simple efficient architecture for sensing active nodes.

Current interest of WSNs includes optimizing the performance of sensor networks for distributed sensing applications [2]. It is well known that QoS is an overused term with various meanings and perspectives. Different technical communities may perceive and interpret QoS in different ways. This paper investigates the problem of optimizing sensor network that has sensor nodes whose

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 519–525.

active lifetime is significantly shorter than the required lifetime of the sensor network system. Satisfying these QoS requirements implies that fault tolerance mechanisms through redundancy will be used, which may cause the energy of the system to be quickly depleted. This paper proposes a *self-adaptive methodology* that utilizes *source* and *path* redundancy with the intention to efficiently reduce the required energy consumption while maximizing the lifetime of the sensor system.

The rest of the paper is organized as follows: Section II provides the state of the art research that has done in this area. Section III presents a discussion on the proposed *self-adaptive methodology* for optimizing the number of active sensor nodes. In Section IV, we present the simulation results that show the success of the proposed approach. We conclude the paper in Section V.

II. Related work

WSNs usually contains thousands or millions of sensors, which are randomly and densely deployed (typically 10 to 20 sensors per m^2) [1]. Due to high number of node count it is impossible to keep track of each node and it is not feasible to replace each node in case of their failure. Thus the number of nodes more than the necessary amount to cover the area is deployed to cope with the failure problem. This problem of WSNs is referred as redundancy [1]. This arouses the need for the sensor nodes with greater active lifetime to prevent the failure problem. The main goal is to prolong the lifetime of the network, which can be defined in several ways: (1) the time when the first node depletes its battery, (2) the time till a given percentage of the sensors have enough energy to operate, and (3) the time till a given percentage of the region is covered by active sensors [5].

Recently, much research has done in the area of energy saving issues in WSN's. Many proposals are put forward to minimize energy consumption in sensor networks [5, 6, 7]. Various power saving schemes have been proposed not only for the hardware and architectural design, but also for designing the algorithms and protocols at various layers in the network architecture.

In [5], a theory is proposed to save the energy by reducing the range of communication and the amount of data transmitted where as in [8] a model is proposed in which the sensor nodes are forwarded to sleep mode whenever the nodes are not sensing the environment. In addition, each location of the physical environment is kept under the examination with a set of sensor nodes (with different sensing periods allotted) and rest of the sensor nodes goes to the sleep mode. In [6], a particular location in the environment is sensed by only one sensor node at a time. But as a single sensor node is prone to failure, there may be errors in the sensing.

The majority of work till now considered sensor networks to be entirely immobile. In [5], a new concept of moving the BS to a position at which the distance between the active sensor nodes and BS becomes optimum has been introduced. This paper proposes a novel approach of *self-adaptive system* in which the BS is kept mobile as well as majority of the sensor nodes are forwarded to the sleeping mode by optimizing the number of active sensor nodes. However, for distance optimization, we follow the proposed approach of [5].

III. The use of self-adaptive methodology for optimizing the active sensor nodes

Before going to introduce the proposed *self-adaptive methodology*, it is worth mentioning some of our key assumptions. We assume that each node is aware of its location and they are static in nature. The BS, on the other hand, is unaware of its location and can be moved unlike the nodes. In addition to that, the sensing area of the nodes is assumed to be a circle of radius *r* with the center of the circle as the node itself. Finally, we assume that all the nodes are synchronized with their neighboring nodes and can communicate with them. The first step of the proposed method is the calculation of the optimal distance for placing the BS. We consider that the BS moves very fast to the respective optimal location soon after calculating the minimum distance. Once we done with the distance optimization, the next step is to use the proposed method that reduces the number of active nodes to communicate with the BS.

A. Strategy for Moving the Mobile BS

In order to move the BS to the optimum position we consider the same theory as proposed by [5]. With respect to the strategy proposed by [5], we place the BS to the position by determining the distances between all the sensor nodes and BS.

B. Self-adaptive Methodology for Optimizing the Number of Active Nodes

The lifetime of a sensor network has two phases: first is the initializing phase and the second is the sensing phase. In the first phase, the sensor senses their own position and synchronizes themselves with the neighboring sensors [8]. After the initialization phase, the sensors are ready to sense the physical environment. The proposed method introduces the strategy for reducing the number of active nodes which sense the area and communicate with the BS. This act of sensing the environment indicates the self-adaptiveness of the proposed methodology that leads the mobile active sensor nodes to determine their optimum position with respect to the nearest BS.

On the other hand, a global search is performed by the BS to determine which mobile active sensor nodes are within the range of its radius as shown in Figure 1 part (a). Based on the global search, the BS takes the final decision for activating or deactivating the connections with the active mobile sensor nodes. The more specific perspective

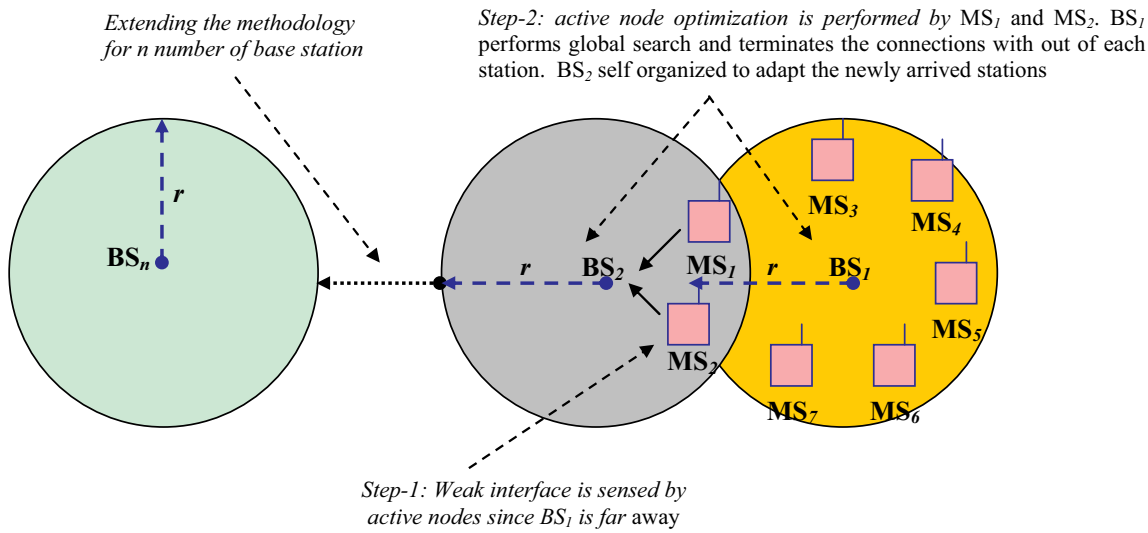


Figure 1 part (a). Self-adaptiveness in base stations and active mobile sensor (MS) nodes for global search and node optimization within the range of *2r*

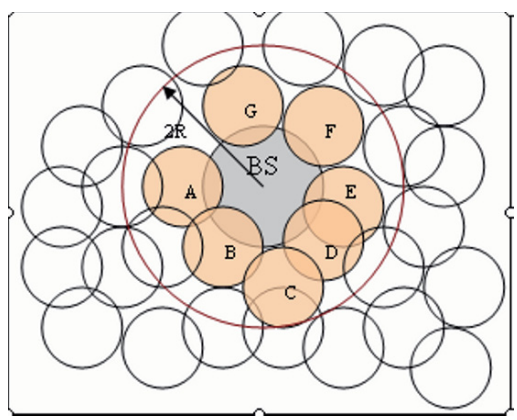


Figure 1 part (b). Behavior of a self-adaptive BS with the active nodes within the range of *2r*

of the proposed methodology can be understood by Figure 1 part (b) that shows the behavior of a self-adaptive BS. As soon as the BS takes its optimum position only those sensor nodes will remain active which are within the range of *2r* of the BS. In Figure 1 part (b), the BS is shown in the middle of a grid and surrounded by the sensor nodes.

In addition, only the sensor nodes which are within the range of 2*r* (Nodes named A, B, C, D, E, F and G) of the BS communicate with the BS. On the other hand, the rest of the nodes which are outside the boundary of the radius move to a passive state for that particular BS. When a node goes to sleep mode, it's sensing, communication and computation components can all be asleep and only a timer needs to work and wake up all components according to its predefined schedule. The BS has the knowledge of the coordinates of sensor nodes which are within the range of 2r in its memory.

In general, when the sensor nodes are activated they synchronize themselves with the BS by broadcasting a message. The sensor node will be responded only if it is in

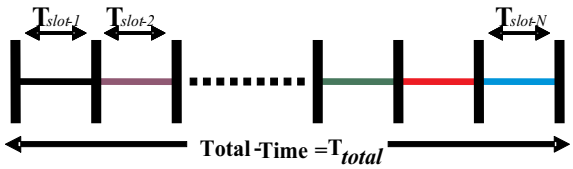


Figure 2. Sensing period of the Active nodes

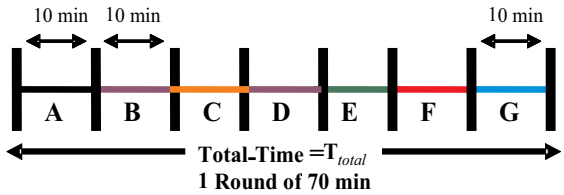


Figure 3. Sensing period for 7 Active nodes

the range of 2r of the BS otherwise it goes to sleep mode. In a given scenario, one may ask that there is a very high possibility of the collision when all the activated sensor nodes broadcast the message to the same BS. In order to solve this problem, the proposed methodology enforces the sensor nodes to activate sequentially. This sequential activation becomes one of the reasons to provide random jitter that assures the activation of nodes. The placement of the BS to the optimum position reduces the distance between the active nodes and the BS which consequently minimizes the communication distance between them.

Thus, the proposed methodology provides these reductions in the distance and the communication cost that become one of the reasons for giving better energy consumption.. In other words, the proposed methodology minimizes the number of active nodes that reduces the total energy consumption. The next step of the proposed method is the division of the sensing period of the remaining active nodes. Previous works has the concept of keeping only one sensor node to be active for sensing a particular region in the environment and so only one node consumes energy and hence the energy is saved [6]. However, one of the common problems with a single node is that it is prone to failure very often which may introduce errors or false alarms. In order to overcome this problem, our proposed method uses the concept of a location which is sensed by more than one node alternately for regular interval of time.

According to our model, the sensing is always done by the active nodes. The active node sensing occurs in the form of cycles of predetermine equal time *Ttotal*. Each cycle (*Ttotal*) is divided into number of time slots *Tslots* with respect to the number of active nodes. This division provides the required time to perform sensing for one active node. This scenario is shown in figure 2. The value

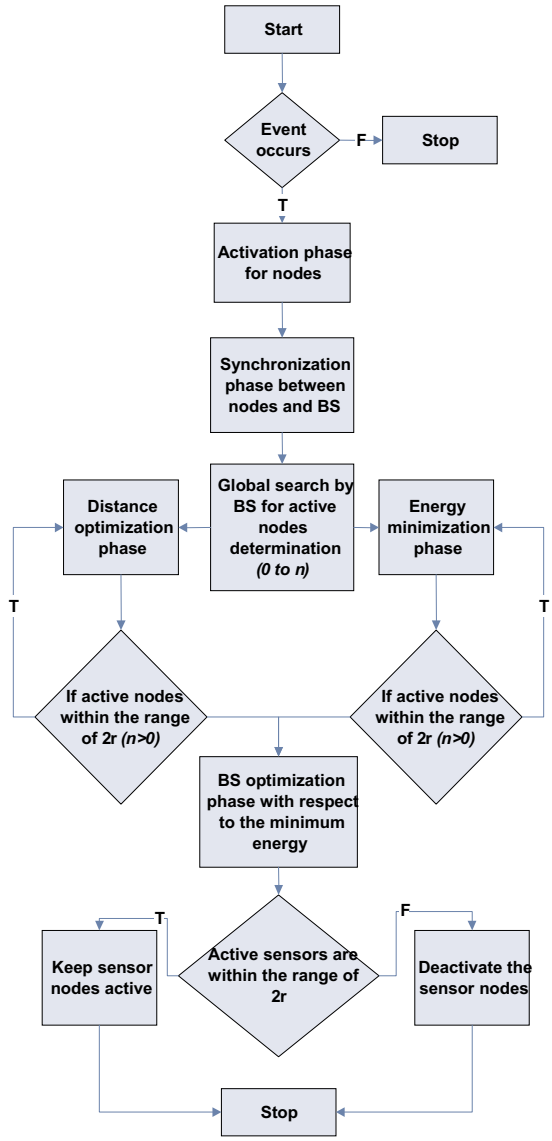


Figure 4. The top-down model for reducing the energy consumption

of *Tslot* depends on the number of optimized active sensor nodes *N*. This relationship can be expressed as:

$$T_{slot} = T_{total} / N \quad (1)$$

To understand this concept, let us assume one cycle of the sensing period *Ttotal* of 70 minutes. For our numerical calculation we kept *N* to be 7 active nodes which come under 2r range of BS. The use of 7 active nodes implies that we need to divide this sensing period among the 7

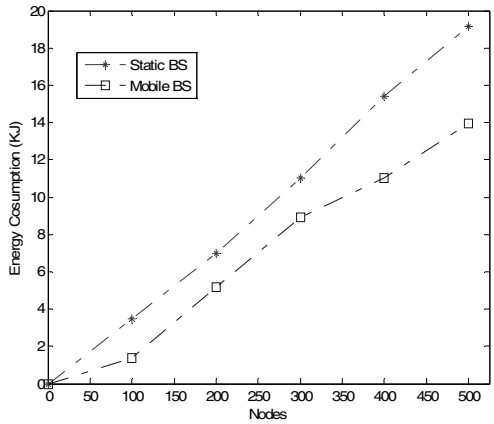


Figure 5. Nodes vs. energy with no. of active-nodes = 500 and BS-radius = 35 mm

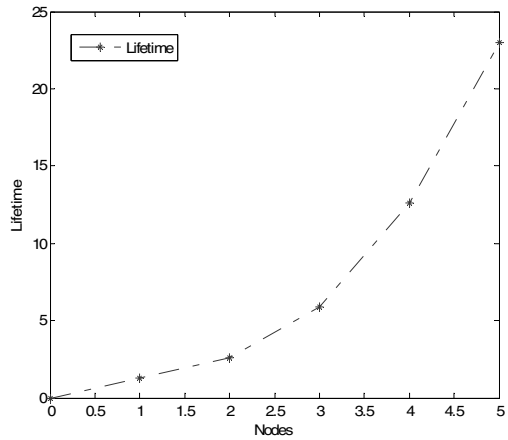


Figure 6. Nodes vs. life time with no. of active-nodes = 5 and BS-radius = 35 mm

sensing nodes which are named as: A, B, C, D, E, F, and G as shown in Figure 3. Taking this scenario into account, each node has to sense for 10 minutes alternately in one cycle of sensing period. The orders of the sensing for the nodes are decided by distance from the BS which is calculated by the proposed methodology. The node which is nearest senses first and so on. Consequently, each point in the target environment is covered by at least one working active node and communicating with the BS. Hence 100% sensing coverage is achieved whereas the average lifetime of each sensor nodes increases as the sensing period gets divided among the active nodes.

IV. Performance Analysis of the Proposed Self-Adaptive Methodology

The proposed methodology is modeled and implemented in C language. For the sake of simulation, we assume that the network is divided into small clusters with typically 450 to 500 nodes in each with an appropriate BS. The *self-adaptive methodology* calculates the individual

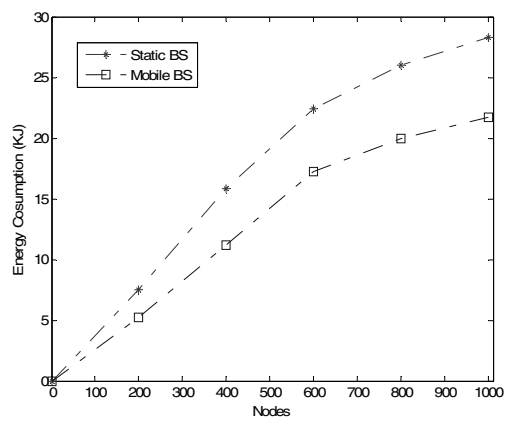


Figure 7. Nodes vs. energy with no. of active-nodes = 1000 and BS-radius = 50 mm

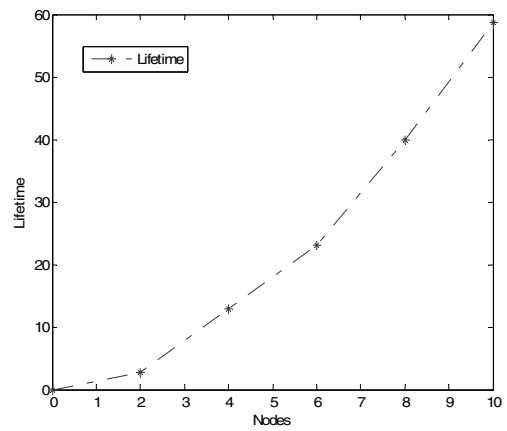


Figure 8. Nodes vs. life time with no. of active-nodes = 10 and BS-radius = 50 mm

distances from the various nodes and the corresponding BS. In addition, we assume that the number of nodes to be 500. After calculating the minimum energy utilized by the nodes, it places the BS to the optimum position.

Furthermore, the proposed *methodology* checks the nodes which are located within the 2r radius of the BS and keeps only these sensor nodes to be active and rest all goes to sleep mode. We assume the value of α to be 2. Note that the value of α can be vary typically from 2 to 5. The proposed top down model can be explained by the flow graph as shown in figure 4.The proposed *self-adaptive methodology* starts working with the occurrence of any event in the environment.

Initially, all the nodes become active and they synchronize themselves with the BS. Our *proposed methodology* starts calculating distance between *n* sensor nodes (where *n* indicates the number of nodes) and the BS until it records minimum energy required to communicate with the BS. Note that the energy consumption is a function of distance between the communicating BS and sensor nodes. The *proposed methodology* places the BS to optimum position where the energy consumed by the network is minimum. Furthermore, our *methodology* verifies the position of each sensor node for decision making. If the targeted nodes are not within the range of 2r of the BS, they are forwarded to sleep mode. As a result, only those sensor nodes are kept active which are within the range of 2r of the BS.

For the sake of performance analysis of the *proposed self-adaptive methodology*, we consider a cluster of network of 500 nodes with the radius of the BS to be 35 mm. The energy required to transmit 1 unit of data is assumed to be 2.5 KJ. The total energy capacity of a node is assumed to be 300 KJ. Note that the node dies as soon as all its energy is consumed. Figure 5 shows the comparison of the energy consumption for a network having static BS with the network having mobile BS. It can be seen in figure 5 that the entire network consumes 3.5 * 10^7 KJ of energy with 100 active nodes when the BS is in the passive state. On the other hand, the total energy consumption is reduced to 1.4 * 10^7 KJ when the BS is moved to the optimum position.

For verifying the success of our proposed model we examined the graph generated by considering 1000 active nodes instead of 500 active nodes and the radius of BS 50mm. Figure 6 explains this scenario. The energy consumed by the entire network with 200 active nodes is 7.5 * 10^7 KJ when the BS is static. Where as this energy reduces to 5.21 * 10^7 KJ after moving the BS to optimum position. One can, therefore, conclude that energy consumption is reduced by more than 45% through our proposed *self-adaptive* methodology compared to other well known methods. Note that as the number of active nodes increases, the difference between the energy consumption for BS static and BS mobile increases significantly which can be seen in figure 5 and 6 respectively which shows the success of our proposed model. Figure 7 and 8 demonstrates the graphs of the lifetime of a particular node when there is more than one active node. According to the proposed methodology, we divide the sensing period among the optimized active nodes. By dividing the sensing period among the nodes, the sensing time for one particular node is reduced significantly. This ensures that the energy required by a particular node to communicate with the BS is reduced and consequently increases the lifetime of the sensor node.

Also, by dividing the sensing period among the optimized active nodes assures that each location in the environment is sensed by at least one active node. This ensures 100% coverage of the environment. Figure 7 shows the lifetime of a sensor node when we considered the radius of BS 35mm and figure 8 shows the lifetime of the sensor node with radius of BS 50mm. We see that in figure 7 the lifetime of one node is 2.6 when the sensing period is divided among 2 active nodes in the network with radius of BS 35mm. The lifetime increases to 2.75 in figure 8 for the same when the radius of BS is 50mm. This is because as the radius of BS increases the sensor node has to communicate for a shorter distance because the effective distance between the sensor node and BS decreases.

V. CONCLUSION

In this paper, we presented a *self-adaptive* system that can be effectively used to optimize the number of active sensor nodes after placing the BS to an optimum position. Once the optimum position of the BS is determined through our proposed top down approach, we can optimize the number of active nodes. The simulation results of the proposed *self-adaptive* methodology demonstrate that the energy consumption of the network is reduced by a very large factor. In addition, our experimental verifications show that the proposed *self-adaptive* methodology provides much longer network life through energy conservation and balancing among sensor nodes. Finally, the simulation results show that the proposed methodology gives comparatively a small communication overhead required to establish a working duty schedule among nodes.

REFERENCES

[1] Hidayet Ozgur Sanli, Hasan Cam, and Xiuzhen Cheng, "An Energy Efficient QoS Protocol for Wireless Sensor Networks," *Proceedings of SCS WMC 2004,*San-Diego, CA.

[2] Jim Kay, "Research Proposal: Optimization of Sensor Networks for the Target Tracking Problem under Power and Reliability Constraints," November 2003.

[3] Ahn Phan Speer, Ing-Ray Chen, "An Optimal path and Source Redundancy for Achieving QoS and Maximizing Lifetime of Query-based Wireless Sensor Networks," *Proceedings of 14th IEEE International Symposium on Modelling, Analysis and Simulation,* mascots, pp. 51-60, 2006.

[4] Kaushal Mittal, Anshu Veda, Bhupendra Kumar Meena "Data Aggregation , Query Processing and Routing in Sensor Networks," IIT Powai, Mumbai, 2004.

[5] D. Vass, Z. Vincze, R. Vida, A. Vidacs, " Energy Efficiency in Wireless Sensor Networks Using Mobile Base Station," *Proc. of 11th Open European Summer Schooland IFIP WG6.6, WG6.4, WG6.9 Workshop (EUNICE 2005),* Colmenarijo, Spain, 6-8 July,2005.

[6] J. Mihaela Cardei, T. Thai, Yingshu Li, Weili Wu, " Energy – Efficient Target Coverage in Wireless Sensor Networks," in *proceedings of the 24th Conference of the IEEE Communications Society(INFOCOM 2005),* March 2005.

[7] C. E. Jones, K. M. Sivalingam, P. Agrawal, and J. C. Chen," A Survey of energy efficient network protocols for Wireless and Mobile Networks," *ACM/Baltzer Wireless Networks,* Vol. 7, Issue 4, pp. 343-358, August 2001.

[8] Ting Yan, Tian He, John A.Stankovic, "Differentiated Surveillance for Sensor Networks," *Proceedings of the 2nd International Conference on Mobile Systems, Applications and Services,* Boston, MA, USA, pp. 270-283 – 2004.

Reducing Malicious Behavior of Mobile Nodes in Ad Hoc Networks

Syed S. Rizvi, Varsha Edla, Saroj Poudyal, and Ravi Nepal
Department of Computer Science & Engineering
University of Bridgeport
Bridgeport, CT 06604 USA
(srizvi, vedla, spoudyal, rnepal)@bridgeport.edu

***Abstract*- A mobile ad-hoc network (MANET) is a kind of self configuring network with mobile routers connected by wireless links. Misbehavior in MANET arises for several reasons such as non-intentionally when a node is faulty or it is overloaded or intentionally when a node is being selfish to save battery, bandwidth, and processing power. This paper presents a new method to stabilize the performance of the network for the active nodes even in the presence of some misbehaving nodes which do not route and forward packets send by the others correctly. This situation becomes even worst when mobile nodes do not rely on any routing infrastructure but relay packets for each other. Specifically, we introduce two extensions to the Dynamic Source Routing (DSR) algorithm [1] to mitigate the effects of routing misbehavior: the *watchdog* and the *pathrater* [2]. Based on the watchdog and the pathrater systems, we propose a reputation trust (RT) algorithm that updates the reputation trust (RT) table of each node with the ranked values of other nodes. Simulation and numerical results show the success of the proposed method by minimizing the malicious behavior of mobile nodes and consequently improving the overall throughput of MANET.**

***Keywords* – Ad hoc networks, mobile nodes, reputation trust, malicious node**

I. INTRODUCTION

In mobile ad-hoc networks, nodes are both routers and terminals. For lack of routing infrastructure, they have to cooperate to communicate. Ad hoc network maximize total network throughput by using all available nodes for routing and forwarding. Hence, the more nodes that participate in packet routing, the greater the aggregate bandwidth, the shorter the possible routing path and smaller the possibility of a network partition is needed. However, a node may misbehave by agreeing to forward packets and then failing to do so, because it is overloaded, selfish, malicious, or broken. An overloaded node lacks the CPU cycles, buffer space or available network bandwidth to forward packets. A selfish node is unwilling to spend battery life, CPU cycles, or available network bandwidth to forward packets not of direct interest to it, even though it expects others to forward packets on its behalf. A malicious node launches a denial of service attack by dropping packets. A broken node might have a software fault that prevents it from forwarding packets.

Without countermeasures, the effects of misbehavior dramatically decrease network performance. Depending on the proportion of misbehaving nodes and their strategies, network throughput decrease, packet loss, denial of service, and can result in network partitioning. These detrimental effects of misbehavior can endanger the entire network. Countermeasures to prevent or to combat misbehavior have been proposed, such as payment schemes for network services, secure routing protocols, intrusion detection, economic incentives and distributed reputation systems to detect and isolate misbehaved nodes. Existing approaches such as economic incentives or secure routing by cryptography alleviate some of the problems, but not all. The main solutions to address this question are secure routing, economic incentives, and detection and reputation systems. Economic incentives such as payment schemes aim at making selfish nodes forward for others despite the power usage and effort this entails. Nodes are paid for forwarding and pay for the forwarding of their own packets by other nodes.

In this paper we accommodate extra facilities in the network to detect and mitigate routing misbehavior. In this way, we can make only minimal changes to the underlying routing algorithm. We introduce two extensions to the DSR [1] to mitigate the effects of routing misbehavior: the *watchdog* and the *pathrater* [2]. The watchdog identifies misbehaving nodes, while the pathrater avoids routing packets through these nodes. When a node forwards a packet, the node's watchdog verifies that the next node in the path also forwards the packet. The watchdog does this by listening *promiscuously* to the next node's transmissions. If the next node does not forward the packet, then it is misbehaving. The pathrater uses this knowledge of misbehaving nodes to choose the network path that is most likely to deliver packets.

Secure routing and economic incentives solve part of the question, but not all. There remains a variety of observable types of misbehavior that they cannot cure easily, such as silent route changes, which may be addressed by detection and reputation systems. They monitor and rate the behavior of other nodes in routing and

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 526–531.

forwarding, such that nodes can respond according to their opinion about other nodes. The goal of a reputation system is to enable nodes to make informed decisions about which nodes to cooperate with or exclude from the network. Reputation systems can be used to cope with any kind of misbehavior as long as it is observable.

The proposed solution could be the use of a self-policing mechanism based on reputation to enable mobile ad-hoc networks to keep functioning despite the presence of misbehaving nodes. The reputation system in all nodes makes them detect misbehavior locally by observation and use of second-hand information. Once a misbehaving node is detected it is automatically isolated from the network.

Another solution to misbehaving nodes is to forward packets only through nodes that share *a priori* trust relationship. A *priori* trust relationships are based on pre-existing relationships built outside of the context of the network (e.g. friendships, companies, and armies). The problems with relying on *a priori* trust-based forwarding are that 1) it requires key distribution, 2) trusted nodes may still be overloaded, 3) trusted nodes may still be broken, 4) trusted nodes may be compromised, and 5) un-trusted nodes may be well behaved. The problem can also be solved for misbehaving nodes by attempting to forestall or isolate these nodes from within the actual routing protocol for the network.

Our paper deals with this problem using the main components in reputation system, the watchdog and pathrater which have been proposed by Marti, Giuli, Lai and Baker [2]. Watchdog is the monitoring part in the network and pathrater is a combined of reputation and response part. The Watchdog detects non-forwarding by overhearing the transmission of the next node. Once misbehavior is detected, the source of the concerned path is informed. For reputation, ratings are kept about every node in the network and the rating of actively used nodes is updated periodically. Node selects routs with the highest average node rating. And hence these two components avoid misbehaving nodes in their routes as a response. This way, network throughput increases as compared to normal DSR.

The response part, in contrast to most other detection and reputation systems does not punish misbehaving nodes that do not cooperate, but rather relieves them of forwarding for others, whereas their messages continue to be forwarded. This way, the malicious nodes are rewarded and reinforced in their behavior. This implies that a watchdog identifies misbehaving nodes and a pathrater helps routing protocols avoid these nodes. When both are used together in a network with moderate mobility, the two techniques increase throughput considerably during extreme mobility, watchdog and pathrater can increase network throughput by 27%, while increasing the overhead transmissions from the standard routing protocol's 12% to 24% [2].

The rest of the paper is organized as follows: Section II presents the state of the art research work. In Section III, we present the proposed Repu-Trust system with the Repu-Trust algorithm. Section IV provides the implementation and the simulation results of the proposed system. Finally, we conclude in Section V.

II. Related Work

There are many approaches that have been proposed to address the misbehavior of node in mobile ad-hoc networks. Few existing solutions are economic incentives, secure routing using cryptography, reputation and response systems, Watchdog and Pathrater, Context Aware Detection etc which aim at reactively detecting misbehavior and proactively isolating misbehaved nodes to prevent further damage. These approaches aim at protecting DSR [1], a reactive routing protocol for mobile ad-hoc networks.

A. Dynamic Source Routing (DSR)

DSR is an on-demand, source routing protocol. Every packet has a route path consisting of the addresses of nodes that have agreed to participate in routing the packet. The protocol is referred to as "on-demand" because route paths are discovered at the time a source sends a packet to a destination for which the source has no path. DSR is mainly divided into 2 main functions route discovery and route maintenance.

Suppose a source node S wishes to communicate with destination node D but does not know any paths to D. S initiates a route discovery by broadcasting a ROUTE REQUEST packet to its neighbors that contains the destination address D. The neighbors in turn append their own addresses to the ROUTE REQUEST packet and rebroadcast it. This process continues until a ROUTE REQUEST packet reaches D. D must now send back a route reply packet to inform S of the discovered route. Since the ROUTE REQUEST packet that reaches D contains a path from S to D, D may choose to use the reverse path to send back the reply or to initiate a new route discovery back to S. Since there can be many routes from a source to a destination, a source may receive multiple route replies from a destination. DSR caches these routes in a route cache for future use.

The second main function in DSR is route maintenance, which handles link breaks. A link break occurs when two nodes on a path are no longer in transmission range. If an intermediate node detects a link break when forwarding a packet to the next node in the route path, it sends back a message to the source notifying it of that link break. The source must try another path or do a route discovery if it does not have another path.

B. Existing Solutions

The main assumption is that nodes are economically rational. Reference [6] proposed incentives to cooperate by means of so called nuglets that serve as a per-hop payment

in every packet in a secure module in each node to encourage forwarding. The secure module is required to ensure the correct number of nuglets is withdrawn or deposited. They propose two models for the payment of packet forwarding, the Packet Purse Model and the Packet Trade Model.

In the Packet Purse Model the sender pays and thus loads the packet with a number of nuglets. Each intermediate node takes one nuglet when it forwards the packet. If there are no nuglets left at an intermediate node, the packet is dropped. If there are nuglets left in the packet once it reaches the destination, the nuglets are lost. This model, however, can lead to the loss of nuglets which have to be re-introduced into the network by a central authority.

C. *Watchdog and Pathrater*

In this section we present the watchdog and the pathrater tools proposed by Marti, Giuli, Lai and Baker [2] for detecting and mitigating routing misbehavior. They call the monitoring part watchdog, and the combined reputation and response part path rater. The Watchdog detects non-forwarding by overhearing the transmission of the next node. Once misbehavior is detected, the source of the concerned path is informed. For reputation, ratings are kept about every node in the network and the rating of actively used nodes is updated periodically.

Nodes select routes with the highest average node rating. The two components enable nodes to avoid misbehaving nodes in their routes as a response. This way, network throughput increases as compared to normal DSR. Michiardi and Molva [3] proposed a Collaborative Reputation (CORE) mechanism that also has a watchdog component for monitoring. A Secure and Objective Reputation-based Incentive (SORI) scheme was proposed by He,Wu, and Khosla [5]. It targets the on-forwarding misbehavior type and uses a Watchdog-like mechanism for monitoring. The reputation system keeps count of the packets forwarded both by and for neighboring nodes. Reputation ratings consist of the ratio of these counts, taking into account the confidence into the rating, proportional to the number of packets requested for forwarding.

D. *Reputation System*

Trust relations are based on evidence related to the previous interactions of entities within a protocol. In reputation system reputation value is a metric for trust. A node with a good reputation means it behaves very well and thus is trustworthy, while nodes with bad reputation are uncooperative and not trustworthy. For each node, reputation rating represents how well a node behaves and trusts rating represents how honest a node is. Reputation value is used to decide whether the node is regular or misbehaved, while trust rating is used to decide whether the node is trustworthy or not, thus the indirect reputation message from the node is accepted or not.

III. PROPOSED REPU-TRUST SYSTEM

To mitigate the misbehavior of Ad Hoc network the Watchdog and Pathrater relies on a method called Reputation System. We have proposed and implemented a local reputation system called “Repu-Trust System” which works in a local network environment. Our System works on top of the existing systems called Watchdog and Pathrater. Before going to depth, a general description of Reputation and Trust is discussed.

A. *Trust and Reputation*

We model trust and reputation as follows:

- Trust is a discrete value of either 1 or 0.
- Reputation is a value between 0 and 1 where 0 represents bad reputation and 1 represent good one.
- Direct trust and reputation are based on direct knowledge or observation.
- Reputation of an indirect node is a product of reputations from intermediate nodes.
- Trust and reputation may not be Reflexive. For example, if node A trusts B, it does not apply that node B also trust A.
- Trust and reputation are usually assumed to be transitive. For example, if node A trust node B and node B knows that node C is to be trusted, then node A can trust C.
- Indirect trust and indirect reputation are calculated based on trust and reputation of their link nodes. For example, if node A trusts node B and node B trusts node C, then A trusts C. On the other hand, if node A does not trust node B, then A will not trust C even if B trusts C.

B. *Repu-Trust System*

The Repu-Trust System consists of essentially five different parts:

1. The detection of misbehavior
2. The calculation and update of reputation values
3. Calculation of Trust Value
4. Broadcasting new Trust and Reputation Value to all local neighbor.

We propose a simple reputation based scheme, called Repu-Trust System, to mitigate misbehavior of nodes and to maximize cooperation. Different from global reputation based schemes, such as CONFIDANT [4] and CORE [3], our solution uses local reputation only. Each node maintains a table with trust and reputation value of all local nodes. The diameter of the network is defined by watchdog. When misbehavior of nodes is found, Repu-Trust System becomes active and starts updating all the table info based on the information provided by its neighbor as well as by Watchdog.

In our scheme, trust in a node is associated with its reputation value. There are only two discrete levels of trust. Symbol ${}_{A}T_{B}$ is used to represent the trustworthiness

of node A on node B. A node A considers another node B either

- To-be-Trusted with T = 1,
- Not-Trusted, with T = 0

A To-be-Trusted node is a well behaved node and nodes trust this with depth of its reputation value. A Not-Trusted node is a misbehaved node and should be avoided and deprived of services. When a new node is entered in the network, it is assumed to be trustworthy with T=1 and Reputation R=1 as well. Its future performance decides its true Trust and Reputation value. It may be a regular or a misbehaved node, depending on its future performance. Every node keeps a reputation table, which associates a reputation value and its trust level with each of its neighbors. It updates the reputation table based on the algorithm defined below. Reputation values R are between a range $0< R < 1$, and there are two thresholds

- T_L below which the node is said to be Not-Trusted.
- T_H above which the node is said to be Trusted.
- So R has a Range ($0 <= T_L < T_H <= 1$)

C. *Implementation of Repu-Trust System*

Each node X maintains trust and reputation value for each of its neighbors in table RT (X). Based on the Repu-Trust algorithm and direct observation, the trust and reputation values are updated. If the reputation value of a neighbor node, for example N, drops below the untrustworthy threshold T_L, then N is considered misbehaved by X. X will notify its neighbors about M's misbehavior by initiating a WARNING message. To prevent false accusations and problems caused by inconsistent reputation values, the WARNING message should be approved by at least three nodes before it can be broadcasted to the local network. Any nodes within one-hop distance of the misbehaved node, M, can sign the WARNING message if, in its reputation table, M's reputation value has also dropped below the untrustworthy threshold T_L.

We make a common assumption that the source node and the destination node are not malicious and the misbehavior if any is only in the intermediate nodes during communication between two trusted source and destination nodes. Consider the scenario in which S sends a message to D using a multi-hop path. As long as the destination is satisfied with the data rate it is achieving and the source is satisfied with retransmission rate, the Repu-Trust system is passive. But as the destination finds that the data rate it is achieving is getting slower and data loss is getting more, it makes an assumption that some node is misbehaving so it initiates a trace process to identify a misbehaved node. This trace process triggers the Repu-Trust System and which in turn activate WatchDog and PathRater and everything becomes functional. Repu-Trust System updates the RT table of each node which comes in the communication path from the additional info it gets from PathRater and Watchdog. Based on it the misbehaved node is found and if needed it is detached from the current communication process for a due time.

D. *Role of Repu-Trust Table*

Repu-Trust uses its RT table to find the Trust and Reputation and also to update it. The RT table of each node is available for view to all the neighbor nodes but they cannot make any adjustments though. On a triggered process every time a node participate in the network protocol, its one-hop neighbors update its reputation value based on Repu-Trust algorithm accordingly. If the participation is positive, then the reputation value is increased, otherwise, the reputation value is decreased.

IV. Simulation And Numerical Results of RT System and Algorithm

Suppose node S wants to calculate the reputation of a node D in its table RT(S) which is not an immediate neighbor of D. I is any intermediate node which follows 2 conditions:

1. Node S trust node I (${}_ST_I = 1$)
2. Node S knows Reputation of node I (R_I is known)

If L is a total amount of data (load) entering node D and L_O is load exiting node D, with total network being W.
Then Reputation of node D is:

- Improved ($R_D < L_O / L$)
- Misbehaved ($R_D > L_O / L$)

Load factor for node D is $\left(\frac{L_O}{W}\right)$ so, the new Reputation is

If improved:

$$R_D = R_D + \left(\frac{L_O}{L}\right) * \left(\frac{L_O}{W}\right) = R_D + \left(\frac{L_O^2}{LW}\right)$$

If Misbehaved:

$$R_D = R_D - \left(\frac{L_O}{L}\right) * \left(\frac{L_O}{W}\right) = R_D - \left(\frac{L_O^2}{LW}\right)$$

If node D has two or more path from node S then the path giving the best Reputation is selected and same value is taken. Based on it, the reputation of D is computed as:

- If $R_D < T_L$ then Trust = 0
- If $R_D \geq T_H$ then Trust = 1

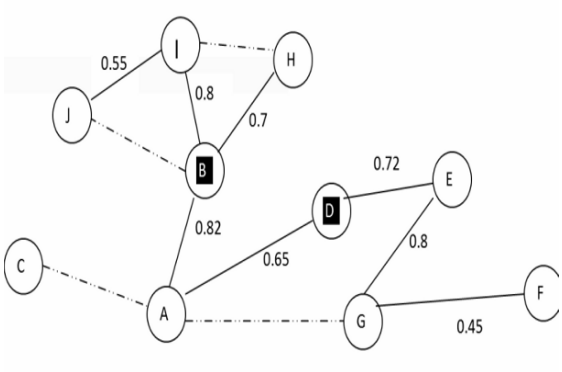


Fig.1 Nodes with trust and reputation

If Trust for a node hence calculated comes to zero the node is found who is misbehaving, so a WARNING message is send to all its neighbor nodes for approval. If it is approved by at least 3 nodes then a broadcast message is send to all the nodes of the network about the misbehavior and the nodes is hence detached from the network for a limited number of time. The duration as a penalty for a misbehaved node depends upon how bad the node was, it is calculates as:

$$\text{Penalty}_{(\text{Dur})} = (T_L - R_D) * T_{(\text{factor})}$$

If the node earlier had a trust of zero and now is behaving good then similar message is send to all the neighbor nodes for approval for goodness. If it is approved by at least 3 nodes then the goodness is broadcasted to whole network which now update the table with that node's trust level set to one.

Fig. 1 shows a network that consists of 10 connected nodes. The dark line shows that the nodes trust each other where as the dotted line shows that nodes do not trust each other. The value given is for the reputation of one node over others. Table I represents the RT table of node *B* before the updates being made where as Table II represents the same RT table after being updated. As shown in Fig. 2, all nodes in the path from source (S) to *D* via *A* are behaving good as their reputations are higher (close to 1.0). This implies that the path raters rating for the path is typically 0.84 as shown in Fig. 4. It should be noted in Fig. 3 that all nodes are well behaved except node *A*, the RT Algorithm hence have dropped the reputation of *A* to 0.4, which results in drop in path rate close to 0.74 as shown in Fig. 4.

TABLE 1
RT TABLE OF NODE *B* BEFORE UPDATE

Node	Trust	Reputation
A	1	0.82
C	1	1
D	1	1
E	1	1
F	1	1
G	1	1
H	1	0.7
I	1	0.8
J	0	0.6

TABLE II
RT TABLE OF NODE *B* AFTER UPDATE

Node	Trust	Reputation
A	1	0.82
C	0	1
D	1	0.53
E	1	0.37
F	1	0.37
G	0	0.37
H	0	0.7
I	1	0.8
J	0	0.6

V. CONCLUSION

In this paper, we presented a new reputation based system to reduce the malicious behavior of mobile nodes in MANET. In addition, we have shown that the use of the

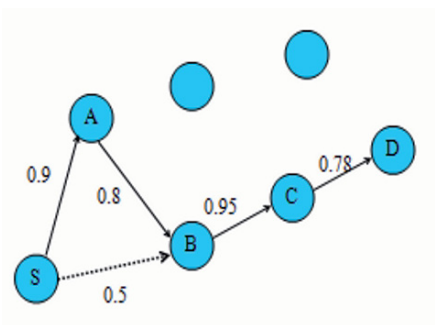


Fig.2. Scenario I

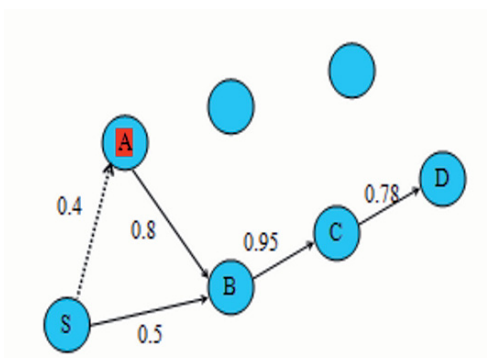


Fig.3. Scenario II

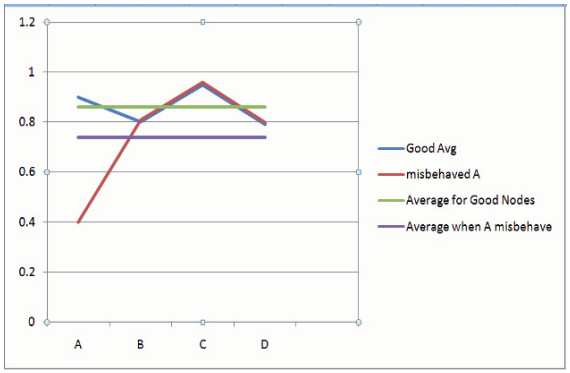


Fig.4. Reputation and Path rating

proposed reputation algorithm to construct the reputation table for each active mobile node can significantly minimize the use of malicious nodes in Ad hoc networks. Based on both simulation and numerical results, one can conclude that even if only single node is misbehaving in a good path, the overall rating of the path drops drastically. As a result, other good nodes try to avoid this malicious behavior in their path. The avoidance results in selecting a good neighbor and isolating the bad nodes from the path as shown in both scenarios I and II.

REFERENCES

[1] K. Paul and D. Westhoff, "Context aware Inferencing to rate a selfish node in DSR based Ad-hoc networks," *in Proceedings of the IEEE Globecom Conference,* Taiwan, 2002.

[2] S. Marti, T. Giuli, K. Lai, and M. Baker, "Mitigating routing Misbehavior in mobile Ad hoc networks," *in Proceedings of MOBICOM 2000,* pp. 255–265, 2000.

[3] P. Michiardi and R. Molva, "CORE: A Collaborative Reputation Mechanism to enforce node Cooperation in mobile Ad hoc networks," *Sixth IFIP conference on security communications, and multimedia (CMS 2002)*, Portoroz, Slovenia, 2002.

[4] S. Buchegger and L. Boudec, "Performance analysis of the confidant protocol," *in Proceedings of the 3rd ACM International Symposium on Mobile Ad Hoc Networking & Computing*, pp: 226-236, 2002.

[5] Q. He, D. Wu, and P. Khosla, "SORI: A secure and objective reputation-based incentive scheme for ad hoc networks," *IEEE Wireless Communications and Networking Conference (WCNC 2004),* Atlanta, GA, USA, March 2004.

[6] Y. Huang, W. Fan, W. Lee, and P. Yu, "Cross-Feature analysis for detecting Ad-hoc routing Anomalies," *in Proceedings of the 23rd International Conference on Distributed Computing Systems (ICDCS 2003), Providence, RI*, pp. 478–487, May 2003.

Application and Evaluation of the LDPC Codes for the Next Generation Communication Systems

Teodor B. Iliev, Georgi V. Hristov, Plamen Z. Zahariev, Mihail P. Iliev
Department of Communication Systems and Technologies, University of Rousse, Rousse 7017, Bulgaria
E-mail: {tiliev, ghristov, pzahariev, miliev}@ccnrgroup.com

Abstract – **Low-Density Parity-Check (LDPC) codes have recently received a lot of attention by their excellent error-correcting capability and have been adopted as an optional error correct coding scheme by many standards. This paper presents, reviews and discussed the implementation of LDPC codes in the upcoming WiMax (IEEE802.16e) and DVB-S2 standards. An analysis and simulation of the LDPC codes has been conducted using various code rates.**

Keywords – **LDPC, DVB-S2, WiMax, Forward error correction.**

I. INTRODUCTION

Low-Density Parity-Check (LDPC) codes and turbo codes are among the known codes nearing the Shannon limit that can achieve very low bit error rates for low signal-to-noise ratio (SNR) applications [1]. When compared to the decoding algorithm of Turbo codes, LDPC decoding algorithm has more parallelization, low implementation complexity, low decoding latency, as well as no error-floors at high SNRs. LDPC codes are considered virtually for all the next generation communication standards.

Mobile WiMax (802.16e) is a broadband wireless solution that enables convergence of mobile and fixed broadband networks through a common wide area broadband radio access technology and flexible network architecture. The performance of Low Density Parity Check (LDPC) codes combined with their relatively simple decoding algorithm make these codes very attractive for the next generation of satellite and radio digital transmission system. LDPC codes integrate the 802.16e standards [2].

The increasing demand of high data rate and reliability in modern communication systems is pushing next-generation standards toward error correction schemes allowing high throughput decoding with performance that is near the Shannon limit.

In particular, an analysis of the LDPC codes for WiFi, WiMax and DVB-S2 has been carried out with the purpose of highlighting the issues and the complexity overhead related to the complete coverage of these standards.

The rest of the paper is organized as follows. Section 2 gives a brief overview of LDPC codes. Section 3 introduces the block LDPC codes. Section 4 describes the standardized LDPC code. Section 5 presents the conducted simulation and the received results. Section 6 concludes the paper.

II. LDPC BASICS

The LDPC codes are linear block codes whose codewords satisfy a set of linear parity-check constraints [1]. These constraints are typically defined by an *m*-by-*n* parity-check matrix *H*, whose *m* rows specify each of the m constraints (the number of parity checks), and *n* represents the length of a codeword. *H* is also characterized by W_r and W_c, which represent the number of 1's in the rows and columns, respectively. A LDPC code can be represented by a bipartite graph, which consists of two types of nodes - Variable Nodes (VN) and Check Nodes (CN). Check node *i* is connected to variable node *j* whenever h_{ij} of H is non-zero. Fig. 1 describes the matrix *H* and the corresponding bipartite graph of a simple LDPC code.

The number of edges on each node is called the node degree. If the node degree is identical for all nodes, the corresponding LDPC code is called regular, otherwise it is called irregular. Note that the communication performance of an irregular LDPC code is known to be generally superior to the communication performance of regular LDPC codes. The degree distribution of the VNs $f\left[d_v^{\max},...,3,2\right]$ gives the fraction of VNs with a certain degree, with $d_v^{\max}$ the maximum variable node degree. The degree distribution of the CNs can be expressed with $g_{d_c}[\max,\max-1]$, where $d_c^{\max}$ is the maximum CN degree, meaning that only CNs with two different degrees occur [2].

A. LDPC decoding process

LDPC codes are decoded iteratively using a message passing algorithm [1]. This algorithm involves exchanging the belief information among the variable nodes and check nodes that are connected by edges in the bipartite graph. Let I_n be the intrinsic information from the received signal, L_n be the reliable information for variable node *n*, $L_{n,m}$ be the information conveyed from variable node *n* to check node *m*, and $E_{n,m}$ be the extrinsic information generated in check node *m* that is passed to variable node *n*. The belief information is updated in an iterative manner and implemented in two phases. In the first phase, the variable nodes send their belief information, $L_{n,m}$, to check nodes connected to them; in the second phase, the check nodes send the updated belief information (new $E_{n,m}$) to the variable nodes connected to them for updating L_n (See Fig. 1).

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 532–536.

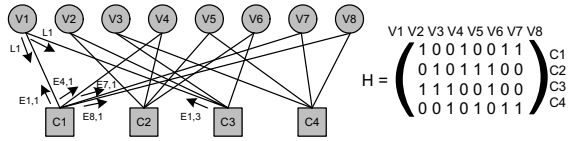


Fig.1 LDPC matrix H and the corresponding bipartite graph

The iteration steps are summarized in Algorithm 1.

Algorithm 1: Min-sum LDPC Decoding Algorithm

1. Initialization: $E_{n,m} = 0, L_n = I_n$
2. VN to CN: $L_{n,m} = L_n - E_{n,m}$
3. Update $E_{n,m} : E_{n,m}^{new} = f(L_{n',m} | n' \in S \subset N(m))$
4. Update $L_n : L_n^{new} = L_{n,m} + E_{n,m}^{new}$
5. Repeat the steps 2,3,4 for NUM iteration times
6. Make a decision of bit n based on the corresponding L_n value

Here, $N(m)$ is the set of variable nodes which are connected with check node m in the bipartite graph. Similarly, $M(n)$ is the set of check nodes which are connected with variable node n. The decoding algorithms differ in how the function f in Step 3 of Algorithm 1 is evaluated. There are three options for the LDPC iterative decoding algorithm: Belief Propagation (BP), λ-min and min-sum algorithms [3]. Although BP and λ-min algorithms show better error correction performance compared to min-sum algorithm, these algorithms require a look-up table for hyperbolic function values, which requires additional memory space. The min-sum algorithm has been selected because of the limited memory size and easy computation patterns. The min-sum algorithm f is shown as follows. Here, $n' \in N(m)$, $n' \neq n$

$$E_{n,m}^{new} = -\left(\prod_{n'} sign(L_{n,m})\right) \times \min_{n'} |L_{n,m}| .$$

As can be seen, the operations in the min-sum LDPC decoding algorithm are limited to addition, subtraction and finding a minimum value.

Theoretically, the LDPC decoding process finishes when all parity-check equations are satisfied. In reality, a predefined number of iterations (NUM) based on SNR is generally used.

B. LDPC matrix partition

A LDPC matrix H has randomly distributed 1's which results in complex data routing and is a major challenge for building a high-performance and low-power LDPC decoder. [4] and [5] show that introduction of some structural regularity in the matrix does not degrade its error correction performance. Moreover the regularity enables partially parallel implementation of LDPC decoders and has been utilized in the IEEE 802.16e standard. Fig. 2 shows the partitioning of H into z-by-z cyclic identity submatrices. Here, I_x represents a cyclic identity matrix with rows shifted cyclically to the right by x positions.

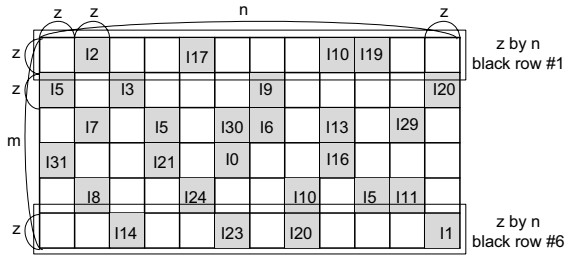


Fig.2 Partitioning of H into z-by-z cyclic identity matrices

This characteristic reduces the routing overhead and has been exploited efficiently in our architecture. Fig. 2 also shows how the n/z of the identity matrices along a row can be grouped to form a block row. So, in essence, the H matrix can also be partitioned into m/z block rows each of size z-by-n.

III. BLOCK LDPC CODES OF WIMAX

A. Description of block LDPC codes

The block irregular LDPC codes have competitive performance and provide flexibility and low encoding/decoding complexity [2]. The entire H matrix is composed of the same style of blocks with different cyclic shifts, which allows structured decoding and reduces decoder implementation complexity. Each base H matrix in block LDPC codes has 24 columns, simplifying the implementation. Having the same number of columns between code rates minimizes the number of different expansion factors that have to be supported. There are four rates supported: 1/2, 2/3, 3/4, and 5/6, and the base H matrixes for these code rates are defined by systematic fundamental LDPC code of M_b by N_b where M_b is the number of rows in the base matrix and N_b is the number of columns in the base matrix. The following base matrixes are specified: 12×24, 8×24, 6×24, and 4×24. The base model matrix is defined for the largest code length (N = 2304) of each code rate. The set of shifts in the base model matrix are used to determine the shift sizes for all other code lengths of the same code rate. Each base model matrix has N_b=24 block columns and M_b block rows. The expansion factor z is equal to N/24 for code length N. The expansion factor varies from 24 to 96 in the increments of 4, yielding codes of different length. For instance, the code with length N = 2304 has the expansion factor z=96 [2]. Thus, each LDPC code in the set of WiMax LDPC codes is defined by a matrix H as

$$H = \begin{bmatrix} P_{1,1} & P_{1,2} & \dots & P_{1,N_b} \\ P_{2,1} & P_{2,2} & \dots & P_{2,N_b} \\ \dots & \dots & \dots & \dots \\ P_{M_b,1} & P_{M_b,2} & \dots & P_{M_b,N_b} \end{bmatrix} = P^{H_b}, \qquad (1)$$

where P_{ij}, is one of a set of z-by-z cyclically right shifted identity matrixes or a z-by-z zero matrix.

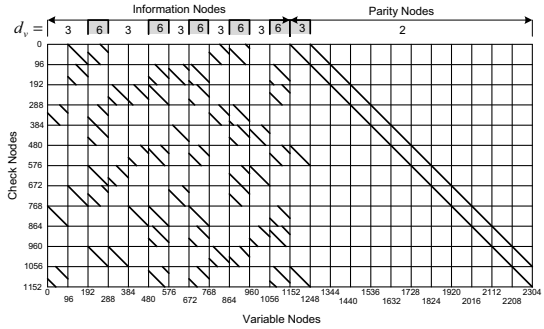


Fig.3 Structure of the parity check matrix for rate1/2 WiMax 802.16e LDPC code (z=96)

Each 1 in the base matrix H_b is replaced by a permuted identity matrix while each 0 in H_b is replaced by a negative value to denote a *z*-by-*z* zero matrix. The codeword length can be calculated by *N*=24.z and ranges from *N*=576 to *N*=2304 bit with a granularity of 96 bit. Figure 3 shows the parity check matrix of the rate 1/2 code for z=96 and thus a codeword length of *N*=2304 bits.

Table 1 summarizes the six code classes with its degree distributions for variable and check nodes. The rate 1/2 code is suitable for layered decoding if the rows are processed in a distinct order. There are two code classes of rate 2/3: code A is highly irregular, code B is semi-regular and allows for layered decoding. The two rate 3/4 code classes differ mainly in the maximum variable node degree to be supported. Rate 5/6 is also provided by one code class. It is obvious that enormous flexibility is necessary to fully support the WiMax 802.16e LDPC code with a one unified architecture:

- 6 different code classes;
- Different VN and CN distribution with $d_v^{\max} = 6$ and $d_c^{\max} = 20$;
- Different sub-matrix sizes from 24×24 to 96×96;
- Different codeword sizes from 576 to 2304 bits;
- Layered and non-layered decoding.

B. Offset min-sum decoding algorithm

Assume binary phase shift keying (BPSK) modulation (a 1 is mapped to -1 and a 0 is mapped to 1) over an additive white Gaussian noise (AWGN) channel. The received values y_n are Gaussian with mean $x_n = \pm 1$ and variance σ^2.

TABLE I
DEGREE DISTRIBUTIONS OF THE WIMAX 802.16E LDPC CODE

Code Rate	f	g
1/2	[2,3,6]={11/24,1/3,5/24}	[6,7]]{2/3,1/3}
2/3A	[2,3,6]={7/24,1/2,5/24}	[10]={1}
2/3B	[2,3,4]={7/24,1/24,2/3}	[10,11]={7/8,1/8}
3/4A	[2,3,4]={5/24,1/24,3/4}	[14,15]={5/6,1/6}
3/4B	[2,3,6]={5/24,1/2,7/24}	[14,15]={1/3,2/3}
5/6	[2,3,4]={3/24,5/12,11/24}	[20]={1}

The reliability messages used in belief propagation (BP) - based offset min-sum algorithm can be computed in two phases: check-node processing and variable-node processing [6]. The two operations are repeated iteratively until the decoding criterion is satisfied. This is also referred to as standard message passing or two-phase message passing (TPMP). For the ith iteration, $Q_{nm}^{(i)}$ is the message from variable node *n* to check node *m*, $R_{nm}^{(i)}$ is the message from check node *m* to variable node *n*, $M(n)$ is the set of the neighbouring check nodes for variable node *n*, and $N(m)$ is the set of the neighbouring variable nodes for check node *m*. The message passing for TPMP is described in the following three steps:

Step 1. Check-node processing: for each *m* and $n \in N(m)$:

$$R_{mn}^{(i)} = \delta_{mn}^{(i)} \max(k_{mn}^{(i)} - \beta, 0), \tag{2}$$

$$k_{mn}^{(i)} = \left| R_{mn}^{(i)} \right| = \min_{n' \in N(m) \setminus n} \left| Q_{n'm}^{(i-1)} \right|, \tag{3}$$

where β is a positive constant and depends on the code parameters [3]. In general, for the irregular codes, we will also apply the correction on variable node messages. The sign of check-node message $R_{nm}^{(i)}$ is defined as:

$$\delta_{mn}^{(i)} = \left(\prod_{n' \in N(m) \setminus n} \operatorname{sgn}\left(Q_{n'm}^{(i-1)} \right) \right).$$

Step 2. Variable-node processing: for each n and $m \in N(n)$,

$$Q_{n'm}^{(i-1)} = L_n^{(0)} + \sum_{m' \in M(n) \setminus m} R_{m'n}^{(i)}, \tag{4}$$

where the log-likelihood ratio of bit *n* is $L_n^{(0)} = y_n$.

Step 3. The decision for final decoding is defined as:

$$P_n = L_n^{(0)} + \sum_{m \in M(n)} R_{mn}^{(i)}. \tag{5}$$

A hard decision is taken by setting $\hat{x}_n = 0$ if $P_n(x_n) \geq 0$, and $\hat{x}_n = 1$ if $P_n(x_n) < 0$. If $\hat{x}H^T = 0$, the decoding process is finished with $\hat{x}_n$ as the decoder output; otherwise the steps from 1 to 3 must be repeated. If the decoding process doesn't end within predefined maximum number of iterations, it_{max}, stop and output an error message flag and proceed to the decoding of the next data frame.

IV. STANDARDIZED LDPC CODES

All current standardized LDPC codes incorporate a common set of features which made feasible implementation possible in the first place: Accumulator-based matrices to allow for linear encoding complexity, structured matrices to be mapped on

partly parallel architectures, and permuted identity sub matrices to ease network implementation.

The DVB-S2 satellite video broadcasting standard [7] was designed for an exceptional error performance at very low SNR ranges (up to FER≤10^{-7} at -2.35dB E_S/N_0). Thus the specified LDPC codes use a large block length of 64800 bit with 11 different code rates ranging from 1/4 to 9/10. This results in large storage requirements for up to 285000 messages and demands high code rate flexibility at the same time to support all specified node degrees.

The current WiMax 802.16e [2] standard features LDPC codes as an optional channel coding scheme. It consists of six different code classes with different VN and CN distributions, spanning four different code rates from 1/2 to 5/6. All six code classes have the same general parity check matrix structure and support 19 codeword sizes, ranging from 576 to 2304 with a granularity of only 96 bit. This codeword size flexibility is the most challenging aspect of this standardized LDPC code family. The interest of the standardization committee in the improvement in throughput and communications performance achievable through layered decoding is evident since an optimal sequence of layers [8] is specified for two of the code classes [9].

The upcoming Wi-Fi 802.11n [10] standard will also feature LDPC codes as an optional channel coding scheme. It utilizes 12 different codes utilizing four code rates from 1/2 to 5/6 for each of the three different codeword sizes of 648, 1248, and 1944 bit. The most complicated issue with this code is the CN and VN flexibility needed to fully support this standard.

Data transmitted across an 802.22 network will make use of adaptive OFDMA with 1024 or 2048 sub-carriers in both the forward and reverse directions. The ability to adapt the number of sub-carriers used will allow 802.22 to be more resilient to interference and other outside influences [11]. Furthermore, by making use of the OFDMA technology developed for WiMax and Mobile WiMax, products based on the final 802.22 standard will be able to be deployed in a much timelier manner.

In terms of modulation, rates from BPSK up to 64QAM supported with the ability to dynamically adapt the method used as channel conditions change. Coding via convolution codes is mandatory with optional support for LDPC and turbo coding. Furthermore, a more advanced method of turbo coding known as Shorted Block Turbo Code (SBTC) is expected to be supported as well. SBTC is based on the Turbo Product Code (TPC) used in 802.16 and is said to provide better parity checking mechanisms than the method used in 802.16e [11, 12].

V. Simulation and Results

To evaluate the proposed LDPC codes, we performed simulations assuming binary phase-shift keying (BPSK) modulation and transmission over the additive white Gaussian noise (AWGN) channel. These simulations have been carried out using MATLAB. For each simulation, a curve showing the bit-error rate (BER) versus E_b/N_0 was computed. The E_b stands for energy per bit and the N_0 stands for the noise power spectral density ratio.

The performance of LDPC codes for WiMax (IEEE 802.16e) and DVB-S2 are illustrated in the figures given below. Simulations were run to determine the performance of these LDPC in AWGN channel with BPSK modulation. In each case the proposed decoding algorithm 1 was used. The decoder stops when either a valid codeword is found or 100 decoding iteration are reached.

Fig. 4 shows the performance of the LDPC code with input frame size of 576 bits and various code rates.

Fig. 5 shows the bit error performance of LDPC code for various code rates and an input frame size of 2304 bits used in 802.16e.

From the conducted simulation study shown on Fig. 4 and Fig. 5 we can conclude that for constant value of BER ($P_B = 10^{-5}$) the necessary SNR for achieving this probability of BER is the most little for larger input frame size.

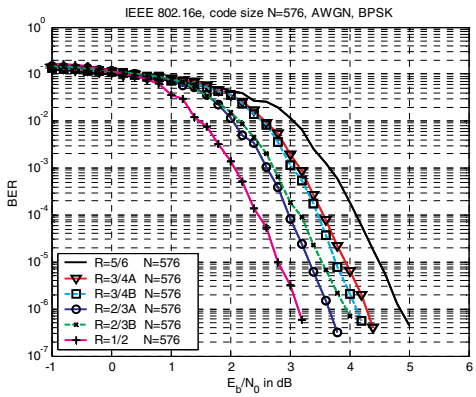


Fig.4. Bit – error performance of the IEEE 802.16e LDPC code for various code rates

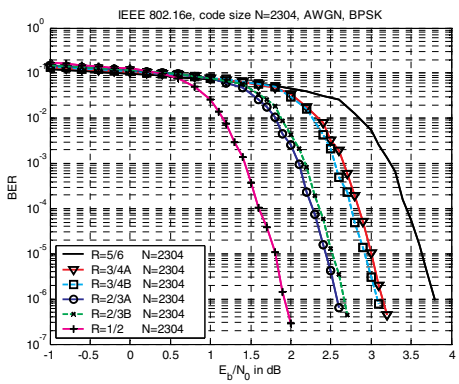


Fig.5. Bit – error performance of the IEEE 802.16e LDPC code for various code rates

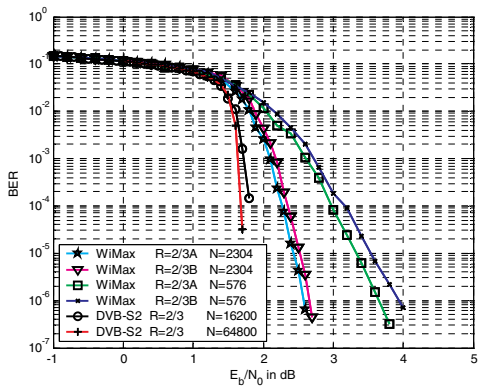


Fig.6. Comparison of the performance of LDPC codes for both 802.16e and DVB-S2 using the same code rate = 2/3

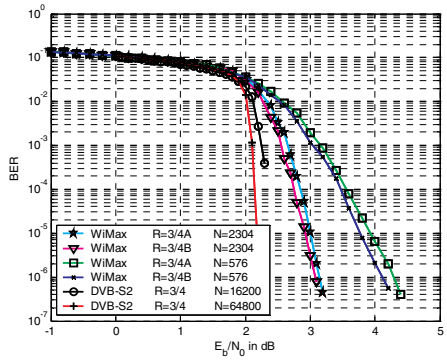


Fig.7. Comparison of the performance of LDPC codes for both 802.16e and DVB-S2 using the same code rate = 3/4

Fig. 6 and Fig. 7 are showing the received simulation dependences of BER for LDPC codes in WiMax and DVB-S2 for two different code rates R=3/4 and R=2/3. By comparing the results obtained, we can observe that DVB-S2 LDPC codes yield a better performance owing to a larger frame size.

VI. Conclusions

LDPC codes were selected for forward error correction in IEEE 802.16e, a reliable broadband metropolitan wireless technology, and DVB-S2, a standard based on DVB-S, which is used for satellite broadcasting. In this paper we have reviewed, presented, validated and discussed the implementation of LDPC codes in both 802.16e and DVB-S2. Secondly, this paper presents an implementation of LDPC codes developed in MATLAB.

The simulation results show that LDPC codes are a powerful error correcting coding technique under SNR environments. It has achieved near Shannon capacity. However, there are many factors need to be considered in the LDPC code design. First, a trade-off between the BER and the number of iterations used in the decoding algorithm need to be made, e.g., more iteration will get lower BER, but the decoding delay is also longer. Secondly, the effect of the frame size on the BER also needs to be considered. Although the LDPC code with larger frame size has better performance, the output delay is also longer. Thirdly, the code rate is another factor that needs to be considered. The higher coding rate needs more bandwidth.

References

[1] R. G. Gallager, "Low density parity check codes," *IRE Trans. Inf. Theory*, vol. IT-8, no. 1, pp. 21–28, Jan. 1962.

[2] IEEE 802.16e. *Air Interface for Fixed and Mobile Broadband Wireless Access Systems*. IEEE P802.16e/D12 Draft, Oct 2005

[3] F. Guilloud; E. Boutillon; J. L. Danger, "λ-min decoding algorithm of regular and irregular LDPC codes," *3rd International Symposium on Turbo Codes & related topics*, September 2003

[4] M. M. Mansour; N. R. Shanbhag, "High-throughput LDPC decoders," *IEEE Transactions on VLSI Systems*, vol. 11, no.6, pp. 976–996, December 2003.

[5] D. E. Hocevar, "A reduced complexity decoder architecture via layered decoding of LDPC codes," *IEEE Workshop on Signal Processing Systems*, pp. 107–112, 2004

[6] T. Iliev, Construction and simulation analysis of Low Density Parity Check Codes. *In: Proceeding of Papers ICEST 2007*, Ohrid, Macedonia, 2007, pp. 35 – 39

[7] European Telecommunications Standards Institude (ETSI). *Digital Video Broadcasting (DVB) Second generation framing structure for broadband satellite applications*; EN 302 307 V1.1.1. www.dvb.org.

[8] M. Rovini, F. Rossi, P. Ciao, N. L'Insalata, and L. Fanucci. Layered Decoding of Non-Layered LDPC Codes. *In Proc. 9th Euromicro Conference on Digital System Design (DSD*), Aug-Sept. 2006

[9] T. Brack, M. Alles, F. Kienle, and N.Wehn. A synthesizable IP Core forWiMax 802.16e LDPC Code Decoding. *In Proc. 2006 Personal Indoor and Radio Communications Conference (PIMRC '06),* Helsinki, Finland, Sept. 2006

[10] IEEE 802.11n. *Wireless LAN Medium Access Control and Physical Layer specifications: Enhancements for Higher Throughput*. IEEE P802.16n/D1.0, Mar 2006.

[11] J. Benko, et. al., "*A PHY/MAC Proposal for IEEE 802.22 WRAN Systems*"

[12] G. Chouinard, "*IEEE 802.22 WRAN System Concept*"

Adjusting the Power Consumption of a Solar Energy Powered Wireless Network Node in Accordance with Weather Forecasts

Thomas Mundt
University of Rostock - Institute of Computer Science - Chair for Information and Communication Services
e-mail: thm@informatik.uni-rostock.de

***Abstract*—We present how weather forecasts estimating the solar radiation for the following days at a given location can be used to adjust the power consumption of a wireless network node. The node is powered by a solar panel which recharges a battery. We discuss two methods to adjust the configuration of the wireless node according to weather forecasts. We are using OLSR routing parameters and hardware adjustments to control the amount of energy needed. One method calculates the available energy, the other uses a naive Bayesian classifier to choose the optimal configuration.**

***Index Terms*—Solar energy, OLSR, Performance optimization, Weather forecasts for solar radiation.**

We thank Christian Schulz, student at the University of Rostock, for performing several simulations and measurements.

I. INTRODUCTION

A. Objectives

This paper discusses results of a practical work. The task was to develop an autonomous network node that runs on solar energy. The node should act as a repeater in a wireless roof top network in a location where no power line is available. All nodes in the network run in IEEE 802.11 ad hoc mode. OLSR is being used as routing protocol.

Solar panels and rechargeable batteries are expensive. The goal was to minimize costs and maximize reliability. Money could be saved by working with smaller solar panels and smaller batteries. The node is powered by a battery which is recharged by a solar panel. The energy which is available to the network node is converted from solar radiation into electrical power. Solar radiation depends very much on the weather, the location of the node, and the season of the year. For this reason we have developed a method to reduce power outages of an autonomous wireless network node. We are using weather forecasts which provide the estimated solar radiation for a given location up to seven days in advance.

We are taking several parameters of the battery, the solar panel, and the wireless node into account to find the best compromise between the nodes normal operation and the need to survive rainy days without power outage. We are using an algorithm which calculates the Bayesian probability of a power outage during the following days. The optimization has the following goals:

- Maximization of the system performance in terms of traffic rate and probably transmission range.
- Minimization of system outtages.
- Minimization of cost for solar panel and recharchable battery.

Controlling the energy consumption is possible by reducing some routing parameters such as "willingness" of OLSR, by adjusting the CPU clock, or by disabling certain hardware components of the node.

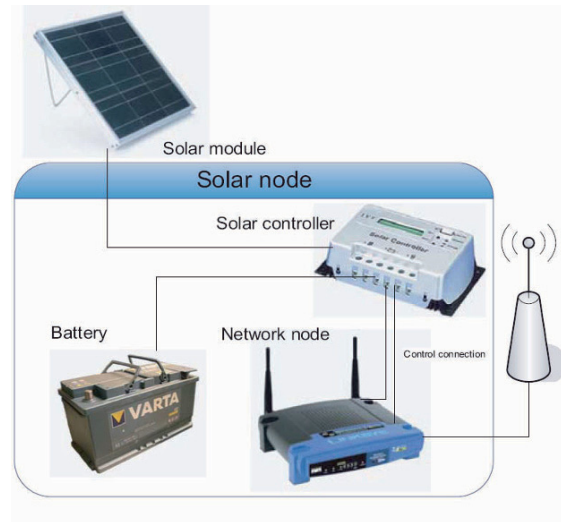


Fig. 1: System overview.

B. Structure of this paper

The remainder of this paper is organized as follows: Section II explains the technological basis. Energy sources such as batteries and solar cells are discussed there as well as possibilities for energy saving, especially through re-configuration of routing and hardware parameters. Section III explains, which calculations and considerations are necessary to reach the optimization goals. An implementation is briefly described in section IV. In section V we present test results for two implemented optimization strategies. Section VI concludes this paper.

II. TECHNOLOGICAL BASIS

To fulfill our task energy demand of the wireless node and auxiliary and energy sources are important. In this section we discuss how solar energy can be used to drive an autonomous network node. We show how the energy balance can be affected on both sides, generation and consumption.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 537–542.

C. Energy sources for autonomous nodes

1) Solar cells: In order to calculate an energy budget for the autonomous node we have to calculate how much energy can be "harvested" by the solar panel. We are using solar panels to recharge a battery during the daylight period. Solar panels consist of many solar cells. These use the photovoltaic effect to convert photons from the sun into electricity using electrons. The energy conversion efficiency is about 15 percent using conventional polymorphic solar cells.

The maximum electric power generated by a solar panel is denoted in Watt peak (Wpx). The panel will generate that power under ideal conditions - which means no clouds, direct sunlight, sun rays strike the panel at a 90 degree angle (perpendicular), no dirt on the panel etc. Standard test conditions are $1000W/m^2$ input, $25°C$ cell temperature, and solar reference spectrum AM1.5 [1] which defines wavelengths and their distribution in the light spectrum. The more peak power is needed the more expensive the panel has to be. Currently the price of polymorphic panels is about 4EUR per Wpx.

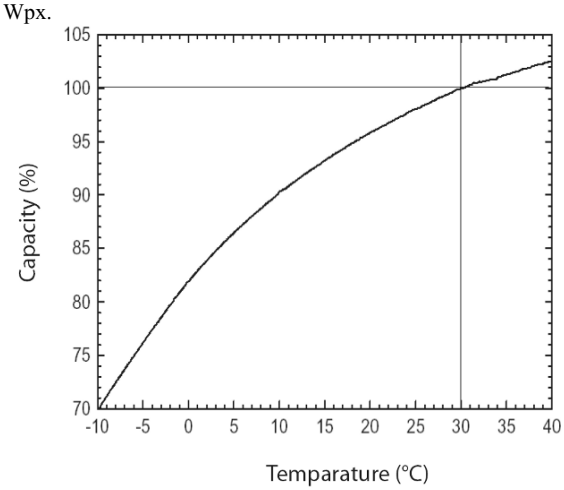


Fig. 2: Capacity depending on temperature.

For testing we used a polymorphic solar panel with a peak power of 55Wpx. Assuming an efficiency of 15% the panel would need around 370W input which requires around $0.37m^2$ effective surface of solar cells under standard test conditions. This fits very well with the panel used for testing with external dimensions of $1\times0.5m$ - some part of the surface is used for wiring and left blank between the cells.

2) Battery technology: There are many different battery technologies available. We concentrate on rechargeable batteries here. Most commonly used are lead acid accumulators, lithium ion batteries, lithium ion polymer batteries, nickel-iron batteries, nickel metal hydride batteries, and nickel cadmium batteries. [2] gives an overview about rechargeable battery technologies.

Lead acid batteries are a good compromise for our purpose as they have a relatively high energy density and a reasonable price. A major disadvantage of lead acid batteries is that they are vulnerable to deep discharge and have a limited number of available load-discharge cycles. An advanced solar controller is able to avoid deep discharge and uses modern loading schemes for recharging such as UIU. An introduction to loading technologies for lead acid batteries can be found at [3] and [4].

B. Energy saving

Generally there are several methods to save energy at an electronic device in general and on a network node in particular. The hardware itself could be optimized to consume as little energy as possible. This method is beyond the scope of this paper and therefore not discussed here.

Other methods are the reduction of traffic through the network node, a decrease of transmission power, and switching off some components of the wireless network node.

There is a variety of routing protocols for mobile ad hoc networks [5], [6], [7]. Some of them such as CEDAR, ScatterWeb, or SAR [8], [9], [10] support energy saving by dynamically changing the paths of packets in order to avoid nodes which have less energy remaining. In the network which is the target for this development we are running OLSR [11] as routing protocol.

1) OLSR routing parameters: OLSR defines a parameter named *willingness* to reflect the current state of the battery and the node's ability to forward packets on behalf of other nodes [12]. The willingness of a node may be set to any integer value from 0 to 7 (WILL NEVER = 0, WILL LOW 1, WILL DEFAULT 3, WILL HIGH 6, WILL ALWAYS 7). A node can dynamically change its willingness as the battery's load changes. A node running on battery should start with WILL DEFAULT and decrease this value step by step to WILL NEVER. The actual energy consumption still depends on the amount of traffic in the network.

Some devices are able to switch off some peripheral components to save energy. On a wireless network node there should be no major unimportant components, but as a means of last resort the transmitting unit might be switched off. Other needless components might be LEDs or unused interfaces.

2) Device modes: We are using an off-the-shelf wireless router Linksys WRT54G as network node [13]. The wireless router is equipped with a Broadcom BCM5325 chipset which includes a 200MHz MIPSEL CPU, 16MB RAM, and 4MB flash memory. This device neither supports any built-in idle or hibernation modes nor has an energy aware hardware or firmware by default. We have retrofitted the router with an RS232 serial interface to connect to the solar controller.

The node's transmitting power can be adjusted between 1 and 250mW - legal requirements allow 100mW EIRP (20dBm) in most countries. Reducing the transmitting power directly affects the number of reachable neighbors and therefore has to be carefully elaborated. The effect on power consumption is marginal. We have tested this by using burst transmissions (*ping -f* to a neighbor node) at 1mW and 100mW. The power input of the entire node decreased from 3.48W (100mW transmitting power, 290mA at 12V) to 3.12W (1mW

transmitting power, 260mA at 12V). The concrete consumption depends on the traffic thus on willingness and general network utilization.

The most significant power saving can be achieved by totally switching off the wireless router's transmitter / receiver unit. This disables the functionality but maintains the possibility to recover later since there is no stand-by mode of the router. The power input decreases from 2.40W (200mA at 12V, transceiver idle) to 1.56W (130mA at 12V, transceiver off).

On the wireless router used as network node we are able to control the CPU clock frequency within about 100MHz and 200MHz (default). This additionally saves a small amount of energy. In experiments we have not noticed any significant decrease in routing performance. Unfortunately, there was no significant decrease in power consumption. The power input dropped from 2.40W (CPU 200MHz, transceiver idle) to 2.28W (CPU 100MHz, transceiver idle). Switching the CPU clock requires a reboot of the wireless router.

Table I summarizes the results and shows that the power consumption can be adapted between 3.48W and 1.44W. To avoid rebooting the router we will only use those modes with a CPU clock of 200MHz (modes 1, 3, 5, and 7).

Mode	CPU clock	Transceiver status	Transmission power	Power input (Current at 12V)
1	200MHz	TX bursts	100mW	3,48W (290mA)
2	100MHz	TX bursts	100mW	3,36W (280mA)
3	200MHz	TX bursts	1mW	3,12W (260mA)
4	100MHz	TX bursts	1mW	3,00W (250mA)
5	200MHz	TX idle	n/a	2,14W (180mA)
6	100MHz	TX idle	n/a	2,04W (170mA)
7	200MHz	TX off	n/a	1,56W (130mA)
8	100MHz	TX off	n/a	1,44W (120mA)

TABLE I
MEASSURED POWER INPUT IN DIFFERENT OPERATING MODES OF A LINKSYS WRT54G WIRELESS ROUTER.

3) Configurations: For simplicity and comparibility of further tests we define four configurations. These are shown in Table II.

Configuration A sets both OLSR willingness and TX power to high performance. Table III shows the resulting energy consumption in our standardized configurations. Through observation of a node in a productive ad hoc network we found that an average node transmits about 20% of the time and idles or listens about 80% of the time. When OLSR willingness is decreased to WILL LOW as in configuration C the transmitter is active for about 5% of the time.

Config.	TX-Idle-ratio	Energy consumption per hour
A	20% transmitting (in mode 1)	8.7kJ (2.41Wh)
B	10% transmitting (in mode 1)	8.2kJ (2.27Wh)
C	5% transmitting (in mode 3)	7.9kJ (2.19Wh)
D	0% transmitting	5.6kJ (1.56Wh)

TABLE III
ENERGY CONSUMPTION DEPENDING ON THE CONFIGURATION AND THE AMOUNT OF TRAFFIC BEING FORWARDED BY A NODE ACCORDING TO ITS OLSR WILLINGNESS.

III. CALCULATIONS

We want to optimize the performance of the wireless ad hoc network. Under ideal conditions every node would run with full power and offer to forward all network packets with highest *willingness*. In section II we have shown several configurations of the wireless node and their respective energy requirements. In this section we discuss the essential idea how to choose the best configurations under several conditions. For this, we are considering a forecast of the solar radiation.

A. Energy budget

1) Solar energy prediction: The part of the solar radiation that reaches the earth's surface depends on the weather, especially clouds. Meteorologists are able to predict the radiation for several days in advance for a given location, but with decreasing accuracy for longer terms. We cannot use those predictions directly, since the efficiency of the solar panel depends on the angle of incidence of the solar rays. To simplify the construction we have decided to use a fixed mounted panel. Figure 3 shows a 7-days-prediction of the global radiation. The prediction was generated at the 5th of June 2006 for the nearby weather observation station Rostock-Warnemuende (at 54◦10*N* 012◦04*E*, the exact position of the solar driven node is 54◦04*N* 012◦07*E*, which is about 12.5km south of the weather station). The chart shows a very sunny week with some clouds predicted for noon of the 8th of June.

To reach an hourly energy sum of 300J per square centimeter (3,000,000J per square meter) a solar radiation of 833W is necessary during the entire hour. The absolute maximum would be 1366W per square meter - which is called the *solar constant* - although it is not constant over time. There are short term and long term variations - the latter is called the *solar cycle*. Both variations are irrelevant for our purpose. The actual maximum for one place depends on its latitude and the daytime.

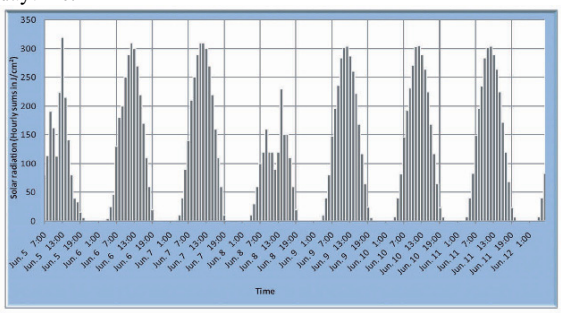


Fig. 3: Solar radiation.

The active surface of the solar panel depends on the angle under which the sun rays strike it. This angle can be calculated by taking time of the day, date, latitude, direction of the panel, and blade angle into account. In order to calculate the angle of incidence we need to know the position of the sun at a given time.

There are arbitrary exact methods such as [14] for this astronomical calculation. We use a simplified method found at [15]. The error is less than 0.1◦

With the fixed blade angle and direction of the solar panel we can easily calculate the angle on which the rays strike the panel like.

This simply leads to an efficiency factor. The predicted radiation energy from the forecast has to be multiplied with that efficency factor (for simplicity we use the full hour when the collection of the hourly sum in the prediction starts as input for the calculation of the sun's position). The fraction of energy that is available for the system furthermore depends on the efficiency of the solar module. To determine this we have performed several calibration measurements.

2) Battery capacity and load: Electrical energy is saved in a rechargeable battery. The general characteristics of a lead accumulator have been explained in section II-A2. Lead accumulators are available with different capacities. The higher the capacity the longer the battery can potentially drive the network node without being recharged by the solar panel.

The current load of the accumulator can be derived from its voltage. Figure 4 shows a typical discharge chart.

Chemical processes destroy the lead accumulators when its voltage falls below a critical value. Therefore solar controllers should shut off the system at this value. The energy available from the accumulator also depends on the age of the accumulator, its temperature, its acid concentration, and the current that flows through the accumulator while being discharged. We need to estimate the available energy. For this reason we have performed several tests. Available input data currently consists of temperature and voltage. The wireless access point used as node consumes approximately 5W. We used a constant power of 5W to discharge the accumulator and constantly monitored the accumulator's voltage. By performing this test several times we collected data to estimate the charging level of the accumulator. We repeated the test several times at temperatures of $20°C$ and $10°C$.

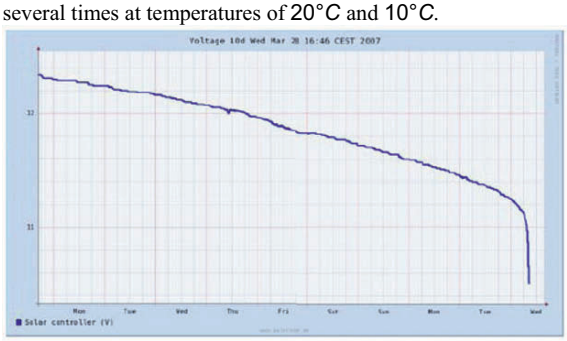


Fig. 4: Discharge chart (voltage over time).

B. Adaptive consumption

The general idea is to adapt the power consumption in order to maximize the node's uptime. We have developed two concepts and two respective prototypes. We call the first approach the calculatory approach and the second one the Bayesian approach.

In the first approach we calculated the remaining life time of the system. For this we used radiation forecast, sun path, current voltage and temperature of the battery, as well as power consumption of the wireless network node. As we are currently able to change the power consumption in four steps, we set four fixed limits for the life time. When the life time falls under a limit the node is powered down one level. When the remaining life time reaches an upper boundary, the node is powered up one level. Calculations are performed every hour.

The second approach uses a Bayesian model to predict the available energy. We used a Bayesian belief in the remaining life time to decide which power consumption level is most suitable for the situation.

In the remaining of this section we will show how we calculated or estimated the life time and how the system performed under several conditions. We compare both methods.

1) Calculatory approach: Basically the remaining life time of the solar driven node is the available energy devided by the current power consumption. In section III-A we have shown how the amount of solar energy available to the system is being predicted. We now focus on the consumption side. The following parameter are available to the system in order to determine the suitable configuration (A to D):

- Battery voltage.
- Temperature of the battery.
- Battery age (currently not considered).
- Time of the day.
- Day of the year.
- Radiation forecast for 7 days.
- Status of the transceiver.
- Size, efficiency, and direction of the solar panel.
- The battery's discharge characteristics (voltage energy relation).

We want to discuss this approach with concrete values. As shown in section II-B2 we can easily set up four levels of energy consumption, although there are unlimited intermediate levels when actual packet rate and power levels between 1mW and 100mW are taken into account.

The car battery (lead acid accumulator) has a rated capacity of 88Ah. The consumption of a fully functional wireless node at 200MHz varies between 240mA (idle) and 290mA (bursts at 100mW transmission power). The solar controller itself additionally consumes 30mA. Consequently the calculated life time with a fully loaded battery varies between 275 hours (11.5 days) and 325 hours (13.5 days). In several tests we have confirmed these figures in reality. In deep sleep mode (configuration C in Table II) the wireless node can "survive" about 22 days without recharging the battery. The node is not functional in this configuration.

As we can foresee at most 7 days in the weather forecast with decreasing certainty, we have chosen 6 days as a reasonable minimum life time to spare. This is about 50 percent of the life time without recharging the battery and with a functional wireless node. With more experiences we could change this later. As mentioned before we consider both energy consumption and energy generation. We calculate the energy budget. When the available energy in the battery (measured as voltage, battery temperature and compared with a table of voltages gathered during several discharge tests) and the

energy being expected to be generated by the solar panel (predicted using the forecast of solar radiation and the angle of incidence) indicates that the solar node will run out of energy earlier than 6 days from the current time we switch down the wireless node one configuration level (from A to B, B to C, or C to D).

We switch the wireless node one configuration up (D to C, C to B, or B to A) when the available energy indicates a spare life time of 8 days. This threshold avoids repeated changes of the consumption configuration.

A short time before the battery is fully discharged the solar controller will disconnect the wireless node in order to avoid deep discharge. This happens at 10.5V battery voltage. As shown in Figure 4 the battery voltage 6 days before the discharge drop is 12.05V (at 20°*C*, we do consider the temperature in the implementation).

2) Bayesian approach: As seen in the preceding Section III-B1 there are a lot of calculations necessary to determine and predict the available energy and the resulting life time.

Some values such as the threshold used to switch the configuration levels up and down are arbitrarily chosen. The network configuration and the traffic rate is very variable and unpredictable as well. Additionally the certainty in the radiation forecasts obviously decreases over time.

Under these circumstances where multiple sources of uncertainty affect the result a stochastic approach with a feedback based mechanism may perform better. To test this theory we have adapted a naive Bayesian classifier [16].

We are using a naive Bayesian classifier $p(O \mid F_1, ..., F_n)$. p represents the likelihood of the outcome O under the conditions $F_1, ..., F_n$, while the outcome O itself represents the four available configurations - see Section II-B2 for a description of those configurations. We choose the configuration O (with values of A, B, C, or D) with the highest probability. We consider the following inputs as features F:

- Temperature of the battery.
- Battery age (currently not considered).
- One separate feature value for every hour of the following 7 days representing the expected energy from the solar panel. These features subsume: Time of the day, Day of the year, Radiation forecast for that day.
- Status of the transceiver.
- Transmission power.
- Routing parameter "willingness" (for OLSR).
- Traffic rate of the wireless node during the last hour.

The following parameters are constant for a single system and have not to be considered:

- Size, efficiency, and direction of the solar panel.
- The battery's discharge characteristics (voltage energy relation).

Every hour we perform a calculation of the Bayesian probabilities. In the probability model we mark the point represented by the current feature vector with the net amount (generation minus consumption) of energy being generated in the system during the next 7 days. This can be either a positive or a negative value. A negative value means we are running out of energy. A positive value means we are not running at the highest possible performance. Both together means, the best value is zero. Currently we are considering positive and negative values as equally insufficient. For this reason we can simply use the absolute value as a criterion. This might be changed in later implementations, for instance in order to overrate survival of the system against performance. For simplicity we use the difference of the current battery voltage and the battery voltage exactly 7 days in the future.

Hence the first 7 days are a training phase for the Bayesian classifier. To shorten this time and to reduce the risk of running out of energy before the self adjusting takes effect we are using a roughly predefined probability model.

The configuration A to D to be chosen is the one with the maximum likelihood for a given feature vector. The probability model will be adjusted permanently during the system's life time.

IV. Implementation

For outdoor tests we have placed the system in an aluminum outdoor case. Figure 5 shows the solar controller and the wireless network node. We are using a modified Linksys WRT54G wireless router for our purposes. This device allows flashing the firmware. We are using an OpenWRT derivative as operating system [17]. We have also retrofitted an RS232 interface [18] to be able to communicate with the solar controller.

Fig. 5: Solar controller.

We have implemented both strategies (calculatory and naïve Bayesian classifier). For the Bayesian model we have used the Bayesian Filtering Library (BFL) [19], which is an open source C++ library for recursive Bayesian estimation. The library is mainly written by the Belgian scientist Klaas Gadeyne, and primarily runs on Linux.

V. Tests

We have tested the system in a real environment. This only allows conclusions about the general fitness of the system, but does not allow comparing the two strategies under equal conditions. For this reason we have set up a simulation environment. It takes more than 10 days until the rechargeable

battery reaches a critical voltage and the solar controller shuts down the system. This is another reason to run a simulation.
The simulation includes:

- Four solar radiation scenarios - one for each season of the year. We collected the necessary real data during two weeks for each scenario by measuring voltage and short circuit current at the solar module (both together can be used to compute the deliverable energy).
- Four corresponding series of radiation forecasts. As the prediction is updated every hour we have recorded the forecast data every hour as well during three weeks – one week to allow the Bayesian classifier to adapt itself, one week for the test, and one additional week to ensure that we have a prediction for the full period of prediction at every time of the simulation including the last day.
- A simulated battery. We have not regarded temperatures and age (number of preceding battery cycles) for the simulation.

We have applied the same simulations with the same simulated data to both algorithms.
The challenging part of this simulation with multivariant data is to define a metric that allows comparing both strategies.
As we want to gain maximum performance with a minimum of energy consumption and a maximum of life time, we decided to measure performance mode of the system and battery voltage every hour.
The following table IV shows a small choice of results.

Time	Calculatory Algorithm		Bayesian Algorithm	
	Config. (A to D)	Battery Voltage	Config. (A to D)	Battery Voltage
Day 01 0000 hours	A	12.0V	B	11.8V
Day 01 1200 hours	A	11.8V	B	11.6V
Day 02 0000 hours	A	11.8V	B	11.6V
Day 02 1200 hours	A	11.8V	B	11.7V
Day 03 0000 hours	B	11.5V	B	11.5V
Day 03 1200 hours	B	11.5V	B	11.5V
Day 04 0000 hours	B	11.4V	B	11.4V
Day 04 1200 hours	C	11.5V	B	11.7V
Day 05 0000 hours	C	11.4V	C	11.5V
Day 05 1200 hours	C	11.5V	C	11.4V
Day 06 0000 hours	C	11.4V	C	11.4V
Day 06 1200 hours	C	11.4V	D	11.4V
Day 07 0000 hours	D	11.2V	D	11.4V
Day 07 1200 hours	D	11.1V	D	11.3V

TABLE IV
EXAMPLE FOR TEST RESULTS. THE TABLE SHOWS THE WINTER SCENARIO. CONFIGURATION A MEANS HIGHEST PERFORMANCE, CONFIGURATION D MEANS TX OFF.

For better comparison we have counted the number of hours for each configuration. Configuration A means highest performance, configuration D means minimum energy consumption.
For better comparison we have calculated a quality indicator as follows. Every hour in configuration A gives 3 points, in B 2 points, in C 1 point and -10 points in configuration D. The weights are arbitrarily chosen and do not represent a classification.

Config	Calculatory Algorithm				Bayesian Algorithm			
	Su	Fa	Wi	Sp	Su	Fa	Wi	Sp
A	90	35	0	30	85	30	19	22
B	39	105	51	71	75	110	80	106
C	39	23	69	31	8	30	28	28
D	0	5	48	36	0	8	41	12
Quality indicator	387	288	-309	-97	413	260	-165	186

TABLE V
NUMBER OF HOURS PER CONFIGURATION. THE TABLE SHOWS HOW MANY HOURS THE WIRELESS NODE REMAINS IN EACH CONFIGURATION DURING THE FOUR SEASONS (TEST SCENARIOS).

REFERENCES

[1] J. Randall and J. Jacot, "Is AM1. 5 applicable in practice? Modeling eight photovoltaic materials with respect to light intensity and two spectra," *Renewable Energy*, vol. 28, no. 12, pp. 1851–1864, 2003.
[2] D. Linden, *Handbook of batteries*. Elsevier Science, 1995, vol. 36, no. 4.
[3] M. Anderson and D. Carr, "Battery energy storage technologies," *Proceedings of the IEEE*, vol. 81, no. 3, pp. 475–479, 1993.
[4] H. Kiehne, *Battery Technology Handbook*. Marcel Dekker, 2003.
[5] C. Perkins and E. Royer, "Ad-hoc on-demand distance vector routing," *Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications*, vol. 2, pp. 90–100, 1999.
[6] J. Broch, D. Maltz, D. Johnson, Y. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols," in *Proceedings of the 4th annual ACM/IEEE international conference on Mobile computing and networking*. ACM Press New York, NY, USA, 1998, pp. 85–97.
[7] C. Gomez, D. Garcia, and J. Paradells, "Improving Performance of a Real Ad Hoc Network by Tuning OLSR Parameters," in *Proceedings of 10th IEEE Symposium on Computers and Communications. ISCC 2005.*, 2005, pp. 16–21.
[8] K. Akkaya and M. Younis, "Energy and QoS Aware Routing in Wireless Sensor Networks," *Cluster Computing*, vol. 8, no. 2, pp. 179–188, 2005.
[9] J. Schiller, A. Liers, H. Ritter, R. Winter, and T. Voigt, "ScatterWeb-Low Power Sensor Nodes and Energy Aware Routing," in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences. HICSS'05.*, 2005, pp. 286c–286c.
[10] V. Paruchuri, A. Durresi, and L. Barolli, "Energy aware routing protocol for heterogeneous wireless sensor networks," in *Proceedings of Sixteenth International Workshop on Database and Expert Systems Applications*, 2005, pp. 133–137.
[11] T. Clausen and P. Jacquet, "Optimized Link State Routing Protocol (OLSR)," in *Proceedings of International Symposium on Performance*. IEEE Computer Society, 2004.
[12] A. McCabe, A. Cullen, M. Fredin, and L. Axelsson, "A Power Consumption Study of DSR and OLSR," in *Proceedings of the Military Communications Conference. MILCOM 2005.*, 2005, pp. 1–7.
[13] "Linksys WRT54G," Documentation available at http://www.linksys.com/.
[14] G. F. P. Bretagnon, "Planetary theories in rectangular and spherical variables. vsop87 solutions." *Astronomy and Astrophysics 202*, vol. 202, pp. 309—-315, 1988.
[15] J. Meeus, *Astronomical Algorithms*. Willmann-Bell, Incorporated, 1991.
[16] D. Hand and K. Yu, "Idiot's Bayes: Not So Stupid after All?" *International Statistical Review/Revue Internationale de Statistique*, vol. 69, no. 3, pp. 385–398, 2001.
[17] S. Innes, "Turning a Linksys WRT54G into more than just a Wireless Router," Edith Cowan University, Tech. Rep., 2006.
[18] C. Strangio, "The RS232 Standard," Rapport, CAMI Research Inc., Lexington, Massachusetts, Tech. Rep., 1961.
[19] http://sourceforge.net/projects/libbpfl/.

A System Architecture for SIP/IMS-Based Multimedia Services

Xianghan Zheng, Vladimir Oleshchuk, Hongzhi Jiao
Faculty of Engineering and Science, University of Agder,
P.O.Box 509, NO-4898 Grimstad, Norway
xianghan.zheng@uia.no
vladimir.oleshchuk@uia.no
hongzhi.jiao@uia.no

***Abstract*- The aim of IP Multimedia Subsystem (IMS) is not only to realize the convergence of the fixed or mobile networks at network level, but also to establish a uni-platform for providing services, no matter which access technology is used. However, due to limited protocol support and the other limitation of device capabilities, a majority of legacy devices (e.g. mobile phone, PDA, laptop, etc) today and in the near future lack coherent support for those services. In this paper, we propose a browser-based Thin Client IMS System Architecture to support non-IMS Client devices. The main issues considered in the paper include security, provisioning, identity management, Thin Client Inter-working, IMS Inter-working and device management. We also build the prototype to validate the design concept.**

I. INTRODUCTION

One ambitious aim of IP Multimedia Subsystem (IMS) is to merge the Internet with the cellular world. As an envisioned solution that will provide new multimedia rich communication services by mixing telecom and data on the access independent IP based architecture, IMS will play an important role in future All-IP infrastructure [1].

Thin Client Computing offers the promise of easier-to-maintain computational services with reduced total cost of ownership [2]. In this computing mode, most of the functionality is allocated on the server side, while client device performs only very simple display and query functions. One typical example of this computing mode is using the browser to explore the Internet [3].

Today, many legacy devices (e.g. mobile phone, PDA, laptop, etc) can not be used for accessing IMS services due to limited protocol support and the other limitations of devices capabilities such as energy, computational power, bandwidth, etc. However, many of these devices have been equipped with various kinds of browsers. These devices can become "thin client" using browser-based thin client mode computing.

The goal of this paper is to propose an appropriate architecture that support non-IMS devices (e.g. mobile phone, PDA, laptop, etc) by using browser-based thin client mode computing. The logical gateway developed based on this system architecture should inter-work between the SIP/IMS signaling and HTTP signaling, and generate HTML-based Graphic User Interface (GUI) corresponding to the services.

II. BACKGROUND

A. *Session Initiation Protocol*

Session Initiation Protocol (SIP) is a transaction-oriented, text-based protocol. It inherited the simplicity from HTTP and Simple Mail Transfer Protocol (SMTP) [4]. SIP with respect to future communication system is regarded to be as important as HTTP to the Internet. The Basic SIP specification RFC 3261 defines six basic methods for the session establishment and termination: "REGISTER", "INVITE", "ACK", "OPTION", "CANCEL", "BYE". Besides, it also has some extension Methods, such as "INFO", "UPDATE", "PUBLISH", "SUBSCRIBE", "NOTIFY", "MESSAGE", etc [5].

B. *IP Multimedia Subsystem*

IP Multimedia Subsystem is a set of standards under development by 3rd Generation Partnership Project (3GPP) in partnership with a number of other standards [6]. It uses SIP protocol to setup, maintain and terminate multimedia sessions. IMS is expected to be an important solution to the future All-IP network and infrastructure which will make Internet technologies, such as web, email, instant messaging, presence, and video conferencing available everywhere. Some important elements in IMS are: User equipment (UE), Call Session Control Function (CSCF), Home Subscriber Server (HSS), Application Server (AS).

C. *Browser-based Thin Client Computing*

Fig.1. shows the browsed-based solution of thin client computing mode. Using browsers, e.g. WAP 2.0 mobile browsers, Internet Browsers, etc, a client device can communicate with the Application Server somewhere in the network via WAP/HTTP protocol. This mode computing is supported by a large mount of devices, e.g. mobile phone, PDA, PSP, laptop, and PC, etc.

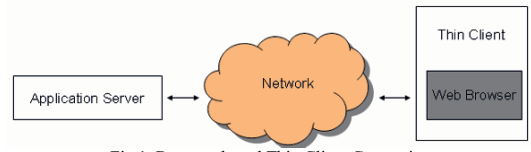


Fig.1. Browser-based Thin Client Computing

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 543–548.

III. Architecture Design

A. System overview

The proposed architecture involves three main characters: Thin Client, Thin Client IMS Gateway (TC-IGW), and the IMS network, as shown in Fig.2.

Thin Client, which can be mobile phone, laptops, PC, etc., uses a web browser to access TC-IGW. The connection between the Thin Client and the TC-IGW could be through either mobile or fixed network. TC-IGW is the key element in this architecture for inter-working between non-IMS and IMS parts. The system utilizes HTTP between Thin Client and TC-IGW, and SIP between TC-IGW and IMS core.

Inside IMS network, there are application servers for handling different IMS services. IMS Application Servers are provided by 3rd party service providers. Typical IMS application servers can be a Media Server, a Presence Server, an Instant Messaging Server, a Push-to-Talk (PoC) Server, etc.

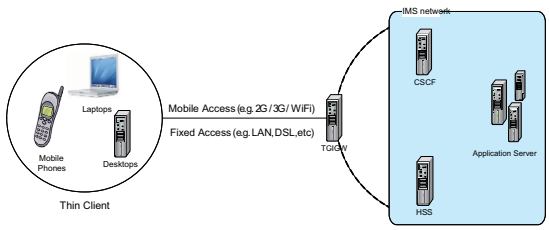


Fig.2. Architecture overview

TC-IGW is the key inter-working unit, acting as a bridge between IMS and non-IMS parts. Note that TC-IGW can also be deployed as an IMS application server with the functionality of an IMS client. Or it can acts as an IMS client outside IMS core network. There are five main components inside a TC-IGW unit (see Fig.3.):

1. Thin Client Inter-working. This part is the portal function between client device and TC-IGW, authenticating thin client users, receiving HTTP messages, generating HTML output.
2. IMS Inter-working. This part is the portal function inter-working between TC-IGW and IMS core, authenticating IMS users, receiving and generating IMS messages.
3. Identity Management. This component stores the user identity and user IMS identity.
4. Device Management. The Device Management function lets the TC-IGW to be aware of the devices connected and their capabilities.
5. O&M Provisioning. This part handles the provisioning configuration, including the functionality of TC-IGW provisioning, identity provisioning. Additionally, this component also provides an administrative interface for the system administrators.

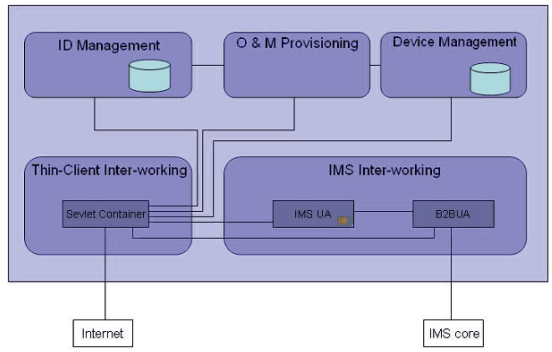


Fig.3. Thin Client IMS Gateway Internal

B. Data Security and User Privacy

We propose to use TLS/SSL to protect the connection between TC-IGW and Internet. The main reason is that most of currently existing browsers (e.g. WAP 2.0 browsers, mobile Internet browsers, desktop browsers, etc.) support (and some of them only support) TLS/SSL encryption mechanism. On the other hand, IPSec can be used to protect the data traffic between TC-IGW and IMS core.

Inside TC-IGW, there is an identity database in the Identity Management component, recording the user private information, e.g. user identity, password, etc. For Internet authentication and authorization, Client user should input the right user identity and password in order to get access.

IMS authentication and authorization depend on the security functionality in the ISIM (IP Multimedia Services Identity Module) application inside the UICC (Universal Integrated Circuit Card) smart card of TC-IGW. ISIM provides the AKA (Authentication and Key Agreement) mechanism that is used to authenticate the subscriber and generate key materials to establish the IP-SEC tunnel between the user equipment and the IMS core.

C. Identity Mapping and Management

TC-IGW holds the whole subscription process for client users, using software logic and its native IP Multimedia Private Identity (IMPI) pre-provisioned by the IMS operator. Three kinds of identities exist: user identity, user IMS identity and private user identity. User identity is the internet-oriented identity, identifying unique user from the Internet point of view. Each user can define his/her own identity, e.g. by using the name of the user, the kind of activity, etc. User IMS identity is the IMS-oriented Identity. Private User Identity is the NAI (Network Access Identifier) format address stored in the UICC smart card inside TC-IGW.

Our identity management solution can be briefly described as following. A user identity is converted to the user IMS identity, defined as, IMPU of the TC-IGW. Then, TC-IGW can use this user IMS identity to communicate with IMS network (see the Fig.4.). TC-IGW has an Internet interface for receiving thin client user identities. The received user identity “user-id” can

be converted into IMPU (SIP URI format) of the TC-IGW as "user-id@imsop.com". This identity is IMS-oriented.

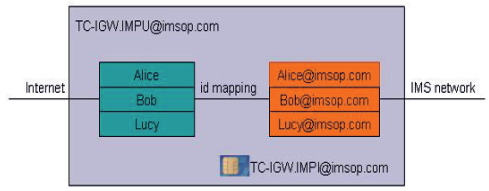


Fig.4. ID Mapping Model

D. Provisioning Configuration

Today, it is very common to configure the mobile devices over the air with SMS or WAP push. We proposed a new Device Management solution defined by OMA Device Management (DM) specification, which is expected to be more promising in the future.

During the device configuration period, Device Management System (DMS) server will first send a notification over SMS to the client device, requesting to initiate a DM session. Inside the SMS message, there is a HTTP(s) address pointing to the download server where configuration file locates. Then the client can be redirected to the OMA Download Server to download and install the configuration package. See the Fig.5 below [7][8].

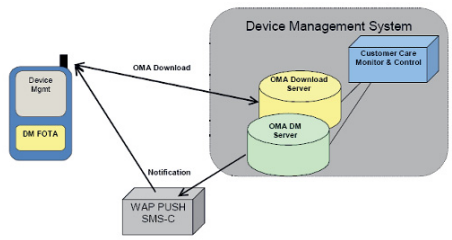


Fig.5. OMA DM Architecture

Auto-configuration is needed each time when TC-IGW initiates. Our proposal is based on UMA (Unlicensed Mobile Access). The operator provides the ISIM inside the UICC card. The ISIM application contains the FQDN (Fully Qualified Domain Name) of the DMS server. When initiation, TC-IGW would fetch the DMS FQDN from the ISIM and initiate a HTTPs session with DMS, download and install the IMS configuration package, as specified before. After auto-provisioning in Step 4 in Fig.6, TC-IGW will register to IMS core.

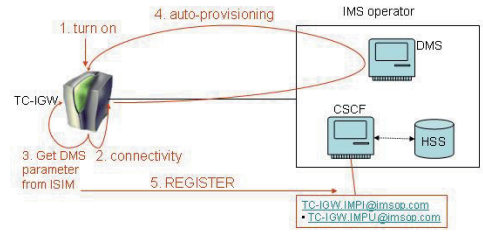


Fig.6. TC-IGW Auto-Configuration

TC-IGW would have to negotiate with the IMS operator for creating a new IMPU for a new user. One potential solution is to use 3GPP GUP (Generic User Profile) protocol specified in 3GPP TS 29 specification [9]. The concept is: TC-IGW uses the SOAP protocol to interact with GUP server in the IMS operator side via Rg interface; inside the IMS operator, GUP server creates / edits the user profile and privacy information via Rp interface, as the Fig.7 shows.

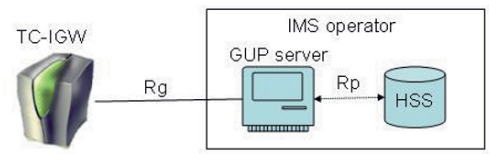


Fig.7. GUP Enabled IMPU Provisioning

TC-IGW also have the interface for administrative to operate and maintain the identity and device database.

E. Thin Client Inter-working

Thin Client inter-working functionality manages the HTTP signal coming from thin clients.

TC-IGW provides Internet interface, generate HTML output for client browsers.

It is better that client browsers receive the updated information automatically so that from end user point of feeling, the message is pushed into the browser. Two categories of interaction solutions exist: Client Pull and Server Push.

Typical Client Pull technologies include HTML time Refresh mechanism, JavaScript Refresh mechanism, and AJAX (Asynchronous JavaScript and XML) mechanism [10]. The first two solutions refresh the whole HTML page to display the renewed information, whilst the later one only updates dynamic element in the Dynamic HTML pages.

Another option is server push. In this solution, client side browser sets up the long-live HTTP connection with the server. Before the connection times out, client browser sends the basic HTTP traffic for maintaining the connection [11]. This solution could save a lot of data traffic; however, it adds burden in server side computing to hold on the connection.

We propose to combine the Client Pull and Server Push technologies together in order to enhance the user experience. TC-IGW is designed to be able to detect the capability of the

client browser by checking the HTTP request message received and intelligently select corresponding handling solution.

F. IMS Inter-working

IMS inter-working is the core IMS functionality unit that manages IMS signaling towards IMS core. There are mainly two sub-components: IMA UA, and B2BUA. IMS UA handles the IMS registration and authentication; B2BUA is responsible to handle the IMS messages, including the generating and parsing of the IMS messages.

IMS UA owns an UICC smart card, inside which USIM or ISIM application exists. TC-IGW has to first register to IMS core in order to become an IMS-enabled client when initiation. After turning on, TC-IGW should firstly fetch the IMPI and P-CSCF FQDN from UICC smart card; then, it could generate and send SIP RIGISTER message to the certain P-CSCF. After that, TC-IGW Registers to the IMS core.

Before the expired registration time, re-registration should be done in order to maintain the IMS connection. Also, de-registration is a must when user logs off.

G. Device Management

Our solution is: inside TC-IGW, there is a device database, recording each connected device and browser capabilities. Based on these different capabilities, TC-IGW intelligently chooses the handling solution, e.g. “PUSH” mechanism, etc.

IV. Use Cases

In the following subsections we demonstrate using of the proposed architecture for several application scenarios.

A. Presence Service

In presence service use case, a user can publish, subscribe, and notify the presence information to a certain user or group of users. Fig.8 shows a presence service scenario. Possible messages between the Thin Client and the Presence Server are:

0. Client Registration.
1. Thin Client updates its status to the Presence Server.
2. One client subscribes the status of another client.
3. The Presence Sever notifies the status to the other client(s).

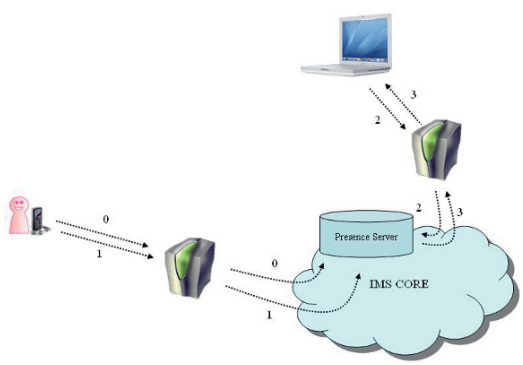


Fig.8. Presence service

B. Instant Messaging

In Instant Messaging use case, a user could send out and receive text and multimedia message to the remote IMS client or similar. Fig. 9 shows an instant messaging scenario. Possible messages between the Thin Client and the Instant Messaging Server are:

0. Client Registration.
1. Thin Client sends a text message and the others receive it.
2. One client asks for setting up an end-to-end multimedia session.
3. Session established, multimedia chat begins.

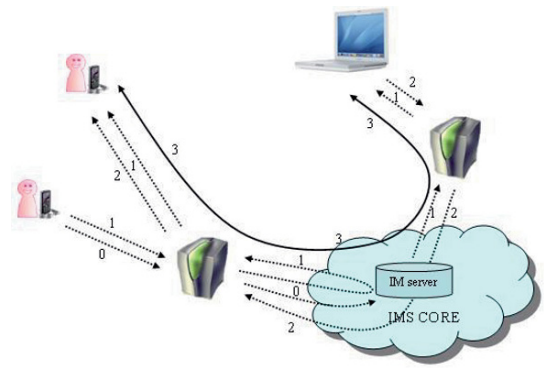


Fig.9. Instant Messaging

C. Content ordering

A user can also order the media content (e.g. music, movie, etc) from the remote IMS application server. See the Fig.10. Possible messages between the Thin Client and Media Server are:

0. Client Registration.
1. Thin Client asks for content list and the AS returns the list.
2. Thin Client asks for downloading the content.
3. Media Server sends the content to the Thin Client.

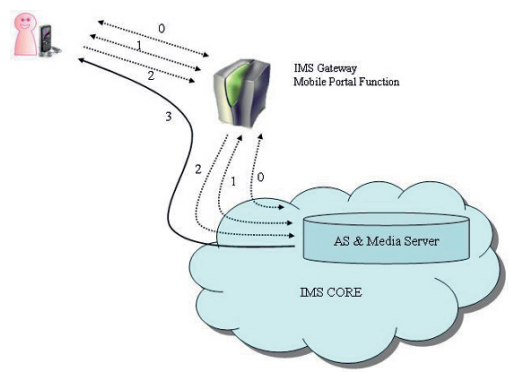


Fig.10. Content Ordering

D. (Re/De-) Registration

Registration must be done before accessing IMS services. Similarly, Re-Registration and De-Registration is needed when the user renews or dismisses the Registration status. See the Fig. 11. Possible messages between the Thin Client and the Registrar are:

0. Client Registration.
1. Client Re-registers Registrar.
2. User De-Registration (e.g. turn off device).

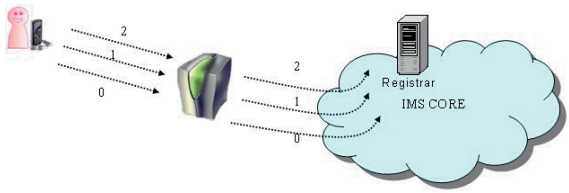


Fig.11. (Re/De-) Registration

V. PROTOTYPE

A. Prototype Environment

We implement the prototype as an IMS application server with three use cases: IMS Registration, Presence service, and Pager mode instant message. The development element and development tools are listed in Table 1.

Table 1
Prototype Implementation Tools

Element	Tools
IMS test bed:	Ericsson SDS SIP Container
IMS test Client	Ericsson SDS Test Agent
DataBase:	MySQL 5.0
Presence Server	Ericsson SDS Presence simulate
Programming Language:	JAVA

B. Structure

Fig. 12 illustrates the structure of the TC-IGW prototype. From message flow point of view, HTTP logic receives and handles the HTTP signals, and then invokes the SIP logic to generate and send out IMS messages. On the other side, SIP logic received IMS message, parse it, and store the useful information into the database, including the identity database for storing user privacy information, and service level database storing specific service data related to the certain use case.

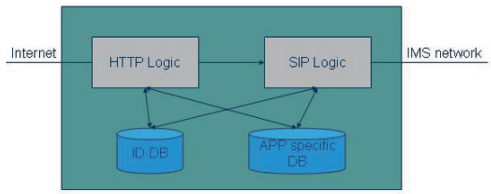


Fig.12. Prototype structure

C. Pre-Condition

We also define some pre-condition for the prototype implementation due to the capability of the Ericsson SDS [12] development tools. The pre-conditions are:

1. We use HTML time refresh mechanism as "PUSH" technology (We set the auto-refresh time as 30 seconds).
2. Security issue is not taken care.
3. Re-Registration and De-Registration is not considered.

VI. EVALUATION

We install IMS core simulator and IMS test Client in Windows 2003 system. Openware WAP 2.0 simulator, Opera Mini mobile browser, SonyEricsson K770i WAP 2.0 browser, and IE 6.0 browser are used as thin client test. We use Ericsson SDS Test Agent as IMS client simulator.

Before testing, we manually configure the user profiles in IMS testing bed environment. We create two users: Alice and Bob, with the corresponding IMPI and also IMPU as alice@ericsson.com and bob@ericsson.com. We assume that Alice uses a browser to access IMS service while Bob uses an IMS client. It is possible to create more user profiles for more tests.

The messages sent out by Alice are instantly forwarded to the Bob's IMS client. However, when receiving IMS message, some delay happens in Alice's client browser. The delay time can be at most 30 seconds. This would probably degrade the user experience.

We find sometimes Bob's IMS client receives duplicated IMS messages from Alice although this message is send only once. The problem might be that the message sent out is lost and retransmitted.

HTML auto-refresh mechanism works well in Openware WAP 2.0 simulator, SonyEricsson WAP 2.0 browser, and IE 6.0; however, it does not work in Operamini simulator. One possible reason is that Opera mini browser is not designed to strictly follow the HTML/xHTML standard.

We have done the functionality testing in three use cases: IMS Registration, Presence Service, and Pager Mode Instant Messaging service. In the first scenario, client browser could register his User IMS identity to IMS core; In the Presence service scenario, client browser could publish its status (e.g. online, busy, away, etc) out to IMS core, subscribe the presence status of the other IMS users. In the Pager Mode Instant Messaging scenario, client browser could send out and receive the text-based messages.

VII. FUTURE WORK

A. Security Issue

TLS/SSL based solution proposed in the system architecture might be not good enough since the solution use password-based access control that may be a weak when users choose passwords that are easy to guess. Approaches based on other access control mechanisms such as, for example, role-based access control, should be considered. Besides, the security between TC-IGW and IMS GUP server is not considered much

in this paper. Additionally, research on the DRM (Digital Right Management) solution (e.g. OMA DRM solution, etc.) should also be done in the future.

B. Portable Solution

TC-IGW can also work as an IMS client together with the functionality of web server. In this solution, TC-IGW should be a relatively powerful IMS device, e.g. advanced smart phone, PDA, laptop, etc, providing the bridge support for a number of non-IMS devices. This may be very useful in reality. For example, one who has non-IMS devices could access the IMS services with the help of his/her friend's IMS device.

C. Extend Use Cases

In the near future, the new release of IMS testing bed, e.g. Ericsson SDS, Open IMS [13], etc, would probably provide more IMS services support. This might enable some fancy IMS services such as session mode Instant Messaging, Push-to-Talk service, etc.

VIII. CONCLUSION

We propose a browser-based Thin Client IMS System Architecture for non-IMS client devices. The system architecture targets on non-IMS client devices. Following the system architecture is the main issues including security, provisioning, identity management, inter-working, and device management. To evaluate the concept and the architecture we implemented three IMS use cases: IMS Registration, Presence service, and page mode Instant Messaging.

REFERENCE

[1] Nisha Rajagopal, Michael Devetsikiotis, "Modeling and Optimization for the Design of IMS networks", Proceeding of the 39th Annual Symposium on Simulation ANSS '06.
[2] Dragos A. Manolescu, George F. Santamarina, "Demonstrations: Java thin clients revisited: an architecture for responsive, live wireless applications", companion of the 17th annual ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications OOPSLA '02, Nov, 2002.
[3] Jerry Golick, "Network computing in the new thin-client age", netWorker, March 1999.
[4] Robert Sparks, "Q focus: session initiation protocol: SIP: basics and beyond", Queue, March, 2007.
[5] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.
[6] Gonzalo Camarillo, Miguel A. Garcia-Martin, "The 3G IP Multimedia Subsystem (IMS) 2nd edition", 2006.
[7] OMA Device Management Protocol, OMA-TS-DM_Protocol-V1_2-20070209-A.
[8] OMA Download Architecture, OMA-Download-ARCH-V1_0-20040625-A
[9] 3GPP TS 29.240, "Generic User Profile (GUP), Stage 3; Network".
[10] Noriko Hanakawa, Nao Ikemiya, "A Web Browser for Ajax Approach with Asynchronous Communication Model", Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence WI '06, Dec, 2006.
[11] Amotz Bar-Noy, Joseph Naor, Baruch Schieber, "Pushing dependent data in clients-providers-servers system", Wireless Network, Sep, 2003.
[12] Ericsson Service Development Studio (SDS) – 4.0. Available: www.ericsson.com/mobilityworld.
[13] Open IMS Playground. Available: http://www.fokus.fraunhofer.de/ims/index.php?lang=en.

Further Improvements to the Kerberos Timed Authentication Protocol

Y. Kirsal and O. Gemikonakli
Middlesex University
The Burroughs, Hendon
London, NW4 4BT, UK

Abstract- **Authentication mechanisms coupled with strong encryption techniques are used for security purposes; however, at times, well-equipped intruders succeed in breaking encryption and compromising data integrity, confidentiality and trust. Kerberos is a widely used computer network authentication protocol which allows individuals communicating over an insecure network to prove their identity to one another in a secure manner. This paper presents an approach that involves temporary interruption to link/server access with a new authentication protocol that delays the decryption with timed authentication. The analysis and verification of authentication properties and results of the designed protocol are presented.**

I. Introduction

Owing to the growing popularity and use of computers and network-based devices, providing privacy and data integrity has become crucial, in order to protect data, resources and systems from attacks and unauthorised access. For the purposes of attack prevention, authentication and access control play a vital role [1]. The basic difference is that the latter one deals with rights and privileges of users who have already gained access into networks, where the former deals with identity verification before access is given.

Kerberos is a commonly used mechanism for authentication purposes. Kerberos utilises symmetric cryptography, to provide authentication for client-server applications. Its implementations allow the introduction of additional algorithms for encryption and check summing. The Kerberos architecture is divided into two core elements, Key Distribution Centre (KDC) and Ticket Granting Service (TGS). The KDC stores authentication information while TGS holds digital tickets for clients and servers of networks for identifying themselves.

Each KDC stores a database of users, servers and secret keys. However, maximum security is essential for the KDCs, since they store secret keys for every user and server on a network. If an attacker can gain administrative access to the KDC, he could have access to the complete resources of Kerberos. If an attacker gains administrative access to a Kerberos client system, he can impersonate the authenticated users of that system. In other words, the authentication service communicates with the TGS and then authenticates the client with a ticket. The TGS receives the ticket from client, checks its validity and replies to the client with a new ticket.

In addition to these, Kerberos exhibits some other vulnerabilities widely reported in literature. Some of these vulnerabilities include among others those known as aided attacks such as replay of old messages, password guessing, SSID sniffing, jamming, masquerading injection, cracking and rouge points or access points, denial of service attacks and session hijacking [2]. There are remarkable efforts to enhance the security capability of this popular authentication mechanism, Kerberos. These efforts are of two categories; public key assistance and the addition of a proxy server [4].

Cryptographic protocols are necessary for secure key distribution to provide suitable keys for authenticated communication. It has also been reported that, the protocols often fail when they are analysed critically [10]. At this point, formal methods emerged for the verification of security protocols. One of the most preferred methods is general purpose verification tools.

The Communicating Sequential Processes (CSP) is an abstract language designed specially for the description of communication patterns of concurrent system components that interact through message passing. The aim of the CSP approach is to reduce questions about security protocols and their properties and ensure that CSP processes satisfy particular specifications [7, 15]. CSP is particularly suitable for describing protocols close to the level we think of them. Schneider states that formalisation of the protocol into CSP exhibits issues and forces design decisions that may not have been distinctly stated in the original protocol description [15]. In addition to CSP, Failure-Divergence Refinement (FDR) is a model-checking tool for state machines, in the theory of concurrency based around CSP. It has an ability to check determinism of a state machine and this is used primarily for checking security properties [14].

Although it is proven that CSP is successful in finding attacks upon a number of protocols, description of a system needs time and substantial experience in order to avoid mistakes. In order to address these concerns, CASPER has been developed [11]. CASPER is a program that automatically produces a CSP description from a more abstract description, thus simplifying the modelling and analysis process. A CASPER script could be divided into two parts: a general part that specifies a model of a system running the protocol in question and a specific part that defines given functions; the parameters of the protocol.

A new protocol description has been designed [7] as a first step in developing a specific authentication protocol. It

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 549–554.

will provide authentication following a previously proposed framework [6]. This protocol has properties of Kerberos and Key-Exchange protocols together with a powerful intruder model. Although the intruder has been given the power to attack, the protocol is successful in preventing replay attacks [8].

Increasing the decryption time of messages for an attacker is another way of attack prevention. Security protocols on distributed systems are time-sensitive. In the analysis of delayed decryption systems, timestamps play an important role. In this paper, a new protocol has been proposed based on the use of timestamps to delay decryption by potential intruders with delay decryption and time authentication properties.

This paper also describes the initial steps involved in developing a specific authentication protocol that has properties of Kerberos and Key Exchange and provides authentication of servers. The proposed framework [6] proposes a security solution to be employed in wireless LANs; an area where the need for high security requirements is beyond doubt. This framework relies on the provisions of IEEE 802.1x standard. It also uses similar infrastructure components as Kerberos but significantly provides for authentication of servers. Throughout this work, CASPER and FDR are used in protocol development and testing.

II. Related Work

In order to provide higher security, various specifications regarding security protocols have been developed and many security protocols have been built around these specifications. Most of these protocols agreed upon a cryptographic key or achieved authentication specifications [1, 9]. These studies indicate that cryptographic protocols are prone to various types of attacks.

Kerberos Authentication Protocol designed as part of project Athena, provides secret key cryptography for secure communications. It is based on Needham-Schroeder Authentication Protocol [12]. It uses key distribution. Clients and servers use digital tickets to identify themselves to the network and secret cryptographic keys for secure communications.

The success of password guessing and replay attacks against Kerberos and weaknesses as a result of Kerberos' requirement of a trusted path have been clearly identified as limitations of Kerberos [2].

Harbitter and Menascé [4] have drawn attention to the performance evaluation of the Kerberos Security Protocol in two different achievements.

Firstly, they used public-key infrastructures *PKINIT, PKCROSS* and *PKTAPP*. In PKINIT, messages are added to change user secret key authentication to public key authentication. It manages secret keys for large number of clients. Nevertheless, it does not address key management of large number of realms. Additionally, as mentioned above, Kerberos uses key distribution and all tickets in its realm are issued by KDC. Since all authentications pass through the KDC, this causes performance bottleneck. At this point, PKTAPP is used for trying to eliminate bottleneck and reduce communication traffic by implementing authentication exchange directly between client and application server. Secondly, in the same study they have proposed the use of proxy servers, *IAKERB* and *Charon* for mobile communication systems. Former one is used as a proxy server, when a client could not establish a direct connection with KDC. Latter one adapts standard Kerberos authentication to a mobile Personal Digital Assistant (PDA) platform. Charon uses Kerberos to establish a trust relationship between a user and a proxy [6].

Nevertheless, as a result, it is possible to say that, although some additional public-key infrastructures have been added to various stages of Kerberos, in terms of server and network capacity, they are suitable for simpler networks and could not work with more than one application server. A proxy server is used to increase encryption process for both client and server. However it produces delays during the transactions of authentication messages between client and server. Recent increase in the speed of wireless networks outperformed proxy servers, leading to insufficient services which resulted in bottle-necks and increased response times [7].

In addition, Kerberos' operation is system and application independent. Kerberos provides a mutual authentication between a client and a server. The Kerberos protocol assumes that initial transactions take place on an open network where clients and servers may not be physically secure and packets travelling on the network can be monitored and even possibly modified. Kerberos is independent of the security features defined in IEEE 802.11.

The framework's three entities (supplicant, authenticator, authentication server) mutually authenticate each other prior to data traffic [6, 8]. It was built on the assumption that none of the parties should be trusted in a wireless local area network communication environment.

Kerberos's mutual authentication uses a technique that involves a shared secret, which works like a password. Many authentication techniques send passwords with no encryption, allowing them to be compromised by an unauthorized party. However, Kerberos solves this problem via encryption. Instead of sending the password, an encrypted key derived from the password is communicated and thus the password is never sent as plain text. This technique can be used to authenticate a client, but can also be used for mutual authentication of a server as well. Once authentication takes place, all further traffic is to be encrypted, allowing even new encryption keys to be communicated securely.

Since, Kerberos KDC software runs on secure hardware, it is assumed that, a roaming user wishes to access an application running on a server somewhere in the network. The user first establishes a secure connection with the Kerberos KDC. After exchanging authentication information, the user, if authorized, gets a "ticket" from TGS which grants access to the desired application. In order to prevent ticket hijacking, Kerberos KDC should be able to

verify that the user who is presenting the ticket is the same user to whom the ticket was issued.

In order to model protocols, the participants in the protocols are modelled as well [3, 15]. In a simple protocol, it is assumed that there are two communicating principals, A and B, and an adversary who is the attacker. The attacker is modelled as having capacity to intercept messages in all directions, modify messages, inject new messages and transmit messages [15].

Apart from those developments in protocol descriptions, the syntax for CASPER scripts is defined as follows [11].

script ::= free-vars-section processes-section
prot-desc-section spec-section
act-var-section [functions-section]
system-section intruder-section

Each section of the CASPER script has different tasks, such as declaration of the agents which are taking part in the protocol, definition of the protocol itself which runs the protocol [8].

- "#Free variables" section declares the type of the free variables and functions used in the definition of the protocol.
- "#Processes" section declares the agents taking part in the protocol, and gives information about their state.
- "#Protocol description" section defines protocol itself, by giving the messages that run the protocol.
- "#Specification" section shows the requirements of the protocol.
- "#Actual variable" section declares the data types used in the system to be checked with FDR.
- "#Functions" section gives definitions for the functions used in the protocol.
- "#System" section defines the system, in terms of the number and types of agents.
- "#Intruder" section gives the identity and initial knowledge of the intruder.

III. Work in Progress

Despite the multiplicity of authentication approaches for improving networks security, forms of attacks have continued to evolve, increasing in number and complexity.

The proposed framework [8] and protocol script [7] provide a design of security solution for wireless local area networks which require high level of security. The requirement for network security is consistent with permitting authorised access to information and services, while preventing unauthorised users from gaining access to and corrupting the network. Since the Kerberos Authentication Protocol is a trusted third party authentication protocol, its paradigms and entities are finalised for the proposed framework [6, 8]. Furthermore, the proposed protocol is a timed model security protocol; it uses timestamps to delay decryption of messages by intruders. An approach has been proposed to shut down external access to an enterprise network for a period of time to enable the distribution of randomly generated keys to users. Renewing keys at various intervals while potential intruders are blocked out would inevitably work against intruders [8].

In the framework, both the program and data containing the credentials of the legitimate entities of a particular wireless LAN environment are installed on each of the entities as well as TGS and KDC. The program adopts the challenge-response paradigm. The interactions between the entities are represented using numbers 1 – 19 [6, 7].

In the previous study [8], the proposed approach was shutting-down external access to an enterprise network for a period of time to enable the distribution of randomly generated keys to users in a relatively secure way. This study is based on the idea that the main threat is from outside and it is relatively easier to control internal sources since the internal sources are known. As mentioned above, keys are renewed at various intervals while the system is shut-down. Also, an analytical model is developed in [8] to evaluate the cost, in terms of performance degradation of the underlying network. Tests carried out for this work and the network with shut-downs revealed that, the protocol in previous study increased the time taken to break the encryption. Additionally, it was shown that, for lightly loaded networks, performance degradation is small, whereas for loaded networks, the performance degradation can be minimised by increasing the number of servers working in parallel. Since, delay decryption mechanism was not used in that study [8], the performance degradation had not been measured

The protocol proposed in this paper is a combination of Kerberos Authentication Protocol and Encrypted Key Exchange Protocol [7] with addition of "delay decryption" property of the Kerberos Authentication Protocol within the designed framework of [8]. Fig.1 shows the proposed system.

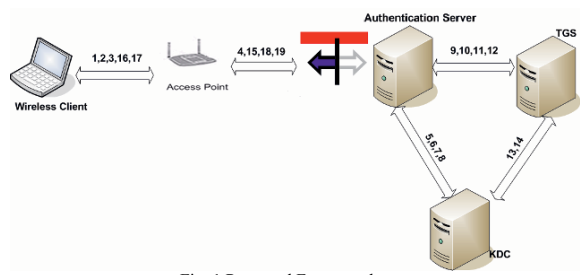


Fig. 1 Proposed Framework

The following script is part of the proposed protocol script written in CASPER:

```
#Processes

INITIATOR(A,S,na) knows SK(A), SPK,
SKey(A), PK(A), passwd(A,B)
RESPONDER(B,S,nb) knows  SK(B), SPK,
SKey(B), passwd(A,B)
SERVER(S,kab) knows PK, SSK(S), SKey,
passwd

#Protocol description

0. -> A : B
[A != B]
1. A -> S : {B}{SKey(A)}
2. S -> A :
{S,A,ts,{kab}{PK(A)},PK(B)%pkb}{SSK(S)}
3. A -> B : {A, ts, na, {kab}{pkb %
PK(B)}}{SK(A)} % v
[A != B]
4. B -> S : {A}{SKey(B)}
5. S -> B : {S, B, PK(A) % pka}{SSK(S)}
[decryptable(v,pka) and
nth(decrypt(v,pka),1)==A \
 and nth(decrypt(v,pka), 2) == now \
 and decryptable(nth(decrypt(v,pka), 3),
PK(A)) \
 and decryptable(nth(decrypt(v,pka), 4),
SK(B))]
<na := nth (decrypt (nth(decrypt(v,pka),
3))) ; \
kab := nth (decrypt (nth(decrypt(v,pka),
4), SK(B)), 1)>
6. B -> A : {nb,na,tb}{kab}
7. A -> B : {nb,ta}{kab}

#Specification

TimedAgreement(A,B,2,[kab])

#Intruder Information

Intruder = Mallory
IntruderKnowledge = {Alice, Bob,
Mallory, Sam, Nm, PK, SPK, SK(Mallory),
SKey(Mallory)}
Guessable = SessionKey
Crackable = SessionKey
Crackable = ServerKey
Crackable = Password
```

In the "#Processes" section, the first parameter of each process (here A, B and S) should represent agent identities used in the "#Protocol description" section. Names are given to the roles played by the different agents. In this protocol, INITIATOR, RESPONDER, and SERVER are chosen. The parenthesized parameters and the variables following the keyword "knows" define the knowledge that the agent is expected to have at the beginning of the protocol run. In the above proposed protocol, the initiator A and S are expected to know his own identity A, S the nonce *na*, *n*b and *ns*, the public key PK, his/her own secret key SK(A), server key SKey(A) and server's public key SPK. Since, actual work of S is designed as SERVER, it is not necessary to define its keys in both INITIATOR and RESPONDER.

In the "#Protocol description" section, the protocol itself is defined by listing the steps in order. These steps will correspond to protocols' messages. Message 0 is used to start the protocol off and tells A, the identity of the agent with whom he should run the protocol. Since the above protocol is timed protocol, timestamps *ta*, *tb* and *ts* are used as variables in the message. Furthermore, the entities of the messages are encrypted with their own public keys and the messages are encrypted with the server's secret key, which is distributed by the server. This increases the time it takes for the intruder to decrypt the message. Especially in message 3, the sender, A creates and sends message, but the receiver, B stores this message in variable *v*, without trying to interpret it. It prefers to wait until it is sure of the authenticity of the communicating agent. In other words, B decrypts this message and performs the appropriate checks only after receiving message 5. Message 3 was encrypted with the inverse of the key receiving in message 5, which B expects to be A's public key.

In the test of message 5 "decryptable", "decrypt" and "nth(_, n)" functions that are provided by CASPER are used [11]. The function "decryptable" takes a message and a key, tests whether the message is encrypted with the inverse of the key. However, the function "decrypt" takes a message and a key and decrypts the message with the key. "nth(_, n)" returns the *n*th field from the message.

Due to delay decryption used, B cannot automatically extract any fields from message 3, so more assignments are needed in the delay decryption model. The assignments are added by using the functions encapsulated between < >. The first assignment assigns the nonce *na* as the third field of message 3, but the message itself, is encrypted with the A's secret key, has to be decrypted using the inverse of this key which is the public key. The second assignment assigns the session key *kab* as the fourth component of message 3 but the first field of the message is encrypted with B's public key.

In addition to these, delay decryption is very sensitive to order of fields in the message. That is to say; if number 1 (written in bold) in the assignment < kab := nth (decrypt (nth(decrypt(v,pka), 4), SK(B)), **1**)> is changed to any other number the output will show a compilation error.

The delay decryption technique that is used in this proposed model is to delay the intruder trying to decrypt a message. Since the proposed protocol is time sensitive, introduction of any delays will help foil the intruder's attempt to launch an attack. In the "#Specification" section

responder B has completed a protocol run with A, implies that A was running the protocol with B within the last 2 time units and there is mutual agreement between them upon the value of session key chosen.

Furthermore, as nonces are also being used for authentication between the agents A and B, the intruder can try to make the first attempt to attack in step 3 when A sends nonce *na* to B. Due to protocol time agreement specification, an agent's chance to attempt to connect will be timed-out by the server because of unsuccessful connection attempts thus preventing the attack.

Additionally, in the "#Intruder Information" section, the intruder's identity and the set of data values that he knows are initially mentioned. The "Intruder knowledge" section holds the identifiers and functions of the protocol that he knows and he can apply any other value to those identifiers and functions.

When the protocol run through CASPER and checked through FDR, and improved throughout the development cycle, due to the strength of the encryption and authentication specifications introduced, there were no attacks found, even when new options "Guessable" and "Crackable" of SessionKey are added to "#Intruder Information" section under the "IntruderKnowledge" option.

IV. Discussions

This paper is concerned with the expression of particular security properties and protocols within CSP and FDR, as well as a compiler tool, CASPER that provides a foundation for analysis and verification.

Additionally, in terms of authentication and authorisation, security aspects of the Kerberos Authentication Protocol are discussed in both wired and wireless networks. Also, this protocol's availability is checked with a security analyser tool CASPER, and the work has proceeded with improvements.

The model, presented above is an addition of a new variant on Kerberos for IEEE 802.11b LANs.

The proposed protocol that has improvements over Kerberos authentication is designed to improve security and minimise possible attacks [7, 8].

As discussed in the previous section, the previous study [8] was about shutting-down external access to an enterprise network for a period of time to enable the distribution of randomly generated keys to users in a relatively secure way. Besides, in terms of performance degradation, an analytical model is designed to evaluate the cost to network performance. The proposed protocol presented in [8] is based on the protocol description of the study presented in [7], where delay decryption mechanism was not used. The results of the analytical model were obtained under these circumstances.

In this paper, the theoretical grounds of a commonly used protocol, Kerberos, its implications, and assuming attackers equipped with possible powers of deduction are presented with inductive capability in CASPER/FDR.

Firstly Kerberos Authentication protocol's capability is checked. After that, new protocol model is designed to minimise possible attacks. Since it minimises possible attacks, new improvements will be introduced on the specifications and description of the protocol.

In order to increase the strength of Kerberos and form a complete security model of wireless networks with all the participants, the developed protocol model will be analysed further. This is done by the validation of "delay decryption" and "timed authentication" encryption of entities and then, encryption of the messages will be tried with different keys as well as passwords. It has been observed that when these changes on the protocol are checked with CASPER, time to break the encryption has significantly increased compared to existing protocol taken as a reference. Nevertheless, it is essential to evaluate the cost of these "delaying decryption" and "timed authentication" properties in terms of performance degradation to the underlying network. This will be performed using an analytical model that will be based on the analytical model of the designed systems defined in [8], developed for the case study. Performance degradation to the underlying network will be highlighted.

V. Conclusions

It becomes highly necessary to improve security authentication and authorisation of networks and computer resources due to the rapid evolution of attacks. This paper is concerned with improving security of the Kerberos Authentication Protocol. A new protocol is derived from the designed and tested protocol [8] to increase the time for an intruder to break an encryption and hence improve the security. The protocol is checked with a security analyser tool, CASPER and has shown to improve security increasing time to break the encryption. This provides a basis to extrapolate the possibilities of what the intruder can achieve with certain knowledge, and where this is achieved.

This paper identifies improved security by restricting externals' access to the system and during distribution of the keys and encryption/decryption of messages is controlled with timed authentication.

It is essential to ensure that increased security will not result in degradation in system performance beyond acceptable levels. For this purpose an analytical modelling approach similar to the one used in [8] can be used to evaluate the system's performance. Results will be obtained by using various decryption times while external links are shut-down. Additionally, these results will be compared with the results that have been obtained from performance degradation of previously designed protocol [8], in order to choose parameters for acceptable levels of security and performance.

References

[1] M. Abadi and R. Needham, "Prudent Engineering for Cryptographic Protocols." *IEEE Transactions on Software Engineering*, vol. 22(1): pp. 6-15, 1996.

[2] S. M. Bellovin, and M. Merritt**,** "Limitations of the Kerberos Authentication System"**.** *USENIX winter 1991,* pp.253-268. 1991.
[3] A. H. Eneh., O. Gemikonakli and R. Comley, "Security of Electronic Commerce Authentication Protocols in Economically Deprived Communities", The Fifth Security Conference 2006, Las Vegas, Nevada, April 2006, ISBN: 0-9772107-2-3.
[4] A. Harbitter and D. A. Menascé, "A Methodology for Analyzing the Performance of Authentication Protocols"**.** *ACM Transactions on Information and System Security,* vol. 5(4): pp. 458-491, 2002.
[5] C. A. Hoare, "Communication Sequence Process". *Prentice- Hall*, International Englewood Cliffs. New Jersey. 1985.
[6] Y. Kirsal, A. Eneh and O. Gemikonakli, "A Solution to the Problem of Trusted Third Party for IEEE 802.11b Networks". *PGNET2005*, Liverpool UK, pp.333-339, 2005.
[7] Y. Kirsal and O. Gemikonakli, "An Authentication Protocol to Address the Problem of the Trusted 3rd Party Authentication Protocols", *CISSE 2006,* 2006.
[8] Y. Kirsal and O. Gemikonakli, "Frequent Key Renewal under Pseudo-Secure Conditions for Increased Security In Kerberos Authentication And Its Impact On System Performability", *3rd International Conference on Global E-Security,* University of East London (UeL), Docklands, UK, 2007.
[9] G. Lowe**,** "An Attack on the Needham-Schroeder Public-key Authentication Protocol**.**" *Information Processing Letters***.** vol: 56(3), pp. 131-133, 1995.
[10] G. Lowe, "Some New Attacks upon Security Protocols", *Proceedings of 9th IEEE Computer Security Workshops*, pp. 162 – 170, 1996.
[11] G. Lowe, "CASPER: A Compiler for the Analysis of Security Protocols". *Proceedings of the 10th Computer Security Foundation Workshop.* pp.18-30., 1998
[12] A. Mishra and W. A. Arbaugh, "An Initial Security Analysis of the IEEE 802.1X Standard", *White paper*, UMIACS-TR-2002-10, February 2002.
[13] M. R. Needham and M.D. Schroeder, "Using Encryption for Authentication in Large Networks of Computers." *Communication ACM* (21) pp.993–999, 1978.
[14] A. W. Roscoe "CSP and Determinism in Security Modelling". *IEEE Symposium on Security and Privacy.* pp. 114-127, 1995.
[15] S. Schneider. "Verifying authentication protocols with CSP" *10th Computer Security Foundations Workshop,* IEEE. pp. 741-758, 1997.
[16] "Security White Paper Evolution, Requirements, and Options" Available: http://wifiplanet.com/tutorials/articles.php/965471 [Accessed: 27 April 2005] [Last Accessed: 13 September 2007].

Self-Repairing Network in a Dynamic Environment with a Changing Failure Rate

Masahiro Tokumitsu[1] and Yoshiteru Ishida[1,2]

[1]Department of Knowledge-Based Information Engineering,
Toyohashi University of Technology,
Tempaku, Toyohashi, 441-8580 Japan
[2]Intelligent Sensing System Research Center,
Toyohashi University of Technology,
Tempaku, Toyohashi, 441-8580 Japan

***Abstract*—We considered a self-repair network by an autonomous and strategic repair. The network is assumed to be composed of agents with a failure rate. This paper further assumes that the failure rate changes dynamically, hence modelling a dynamic environment. When the failure rate oscillates with a fixed amplitude and cycle, computer simulations indicated that there is a threshold of not only the amplitude but the cycle. When the failure rate changes with the cycle not exceeding the threshold, the strategic repair adapts to the environment and exhibits a reasonable performance.**

I. Introduction

As the scale and complexity of a computer network increases, it would be more difficult to repair the failed computer [1,2], or even to locate them manually by human. A self-repair network has been proposed to deal with the problems. Strategic repairs as well as a fixed repair have been studied. The strategic repair demonstrated an adaptive behaviour in a heterogeneous environment when failure rate changes from a computer to computer. In this paper, we study the adaptive behaviour in a dynamic environment where the failure rate changes dynamically. This paper specifically focuses on a situation when the failure rate oscillates with a fixed amplitude and cycle.

II. A Self-Repair Network

A. A Network Model

A self-repairing network consists in agents capable of repairing other agents connected. A strategic self-repairing network has been studied where each agent determines the next action (whether repair neighbour agents or not) depending on its strategy [3]. In the model, each agent is assumed to fail with a failure rate (λ). The failed agents change their state from normal to abnormal.

Agents are expected to repair other abnormal agents, however, the normal agents must consume their own resources to repair other agents. Thus, each agents face a dilemma to determine whether they cooperate (repair by sacrificing their resources) or defect (not repair and save their resources). Obviously, repair rate must be controlled in an appropriate level.

B. Spatial Prisoner's Dilemma

SPD (Spatial Prisoner's Dilemma) has been studied to investigate when, how, and why the cooperation emerges among selfish agents when they are spatially arranged, hence interactions are limited only to their neighbors [4,5]. Each player placed at each lattice of the two-dimensional lattice. Each player has an action and a strategy, and receives a score. Each player plays PD (Prisoner's Dilemma) with the neighbors, and changes its strategy to the strategy that earns the highest total score among the neighbors.

The SPD is generalized by introducing spatial strategy [6]. Spatial Strategy determines the next addition dependent upon the spatial pattern of actions in the neighbors. The strategy has two actions: C (cooperation) or D (defection).In the model, we match the strategy C is repairing and D is not repairing. Figure 1 shows a schematic diagram of the self-repairing network by SPD.

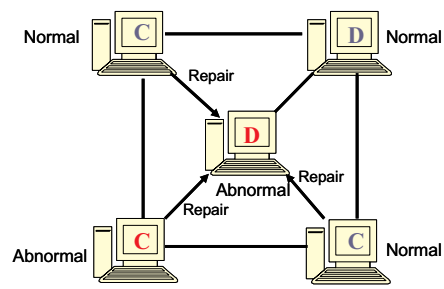


Fig.1 The Self-Repairing Network by SPD in the situation failure rate is high.

C. Self-repair network with Spatial Strategies

Each agent is place in each cell of two-dimensional lattice. They are assumed to be connected with the agents in the neighbour cells. In this study, Moore neighbourhood (eight neighbour agents) is used. Each agent has two states: normal or abnormal. The normal agent can do the task assigned, however, abnormal agent cannot do the task, and even if the

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 555–560.

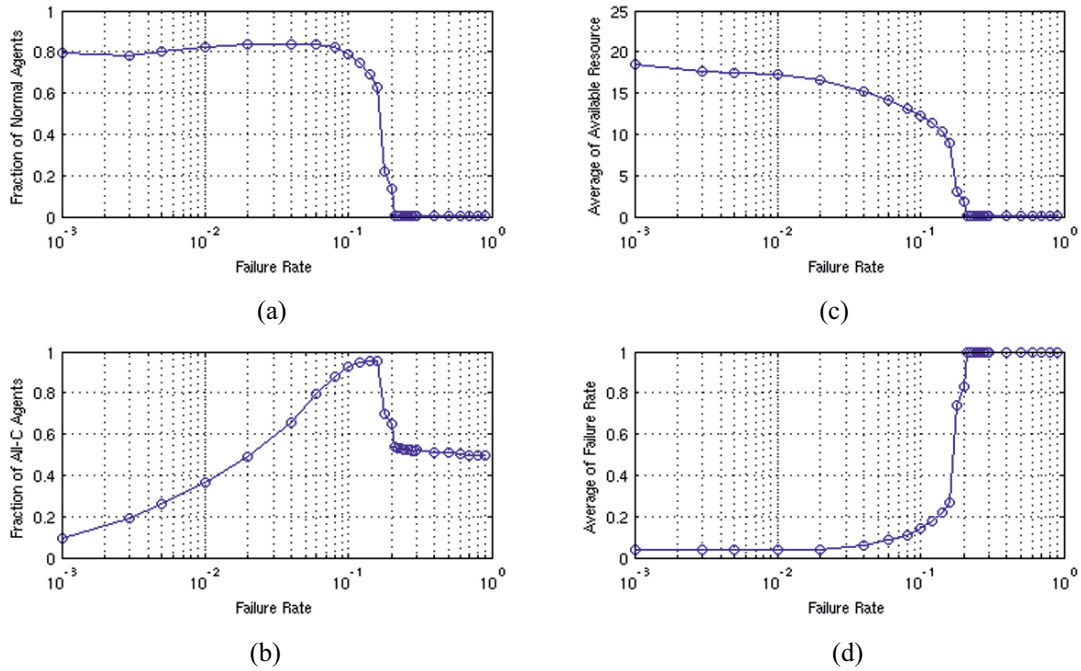


Fig.2 The performance versus failure rate when the failure rate is kept constant. (a) fraction of normal agents (b) fraction of All-C agents (c) the average of available resource per agent (c) the average of failure rate in the network Parameters are as in TABLE 1 and initial configuration with the strategy each agents and 100 failure agents randomly chosen. The failure rate is between 0.001 to 0.90

abnormal agents do an operation, they are likely to fail. Agents are assumed to fail with a dynamic failure rate $\lambda(t)$. The repair will be done by a repair rate α. When the repair is successfully done, the repaired agent becomes normal. Specifically, when the repair is done by an abnormal agent, the repaired agents increase their failure rate by a damage rate δ.

Agents repair neighbour agents at each step. Each normal agent has their available resource R_a. The max available resource is set to $R_{\max}$. Available resource will decrease R_r in repairing another agent. Abnormal agents have no resource.

We apply Spatial Prisoner's Dilemma to the agents. The agents decide their action based on their Spatial Strategy. Spatial strategy determines the next action based on the spatial configuration of the current actions of neighbor agents. We apply two strategies All-D or All-C for the agents. All-D strategy does nothing other agent, All-C repair. The agents change their action C or D by an action error rate (μ). The action error is introduced to prevent the network stuck at the local minima. The agents update own strategy after r (called strategy update cycle) steps from previous change. The next strategy will be chosen from the strategy that earned the highest score among neighbor agents. The total available resource is updated in each step by adding a fixed resource and subtracting consumed resource from the current resource value. TABLE 1 lists simulation parameters used in this study.

TABLE 1

LIST OF PARAMETERS FOR SIMULATIONS

Name	Description	Value
$L \times L$	Size of the space	50 x 50
N	Number of the unit	2500
$N_f(0)$	Initial number of the abnormal agents	100
α	Repair rate	0.1
δ	Damage rate	0.1
r	Strategy update cycle	100
$R_{\max}$	Maximum resource	25
R_r	Resource used for repair	1
μ	Action error rate	0.1
T	Cycle of the failure rate function	10-10000
A	Amplitude of the failure rate function	0.001-0.9

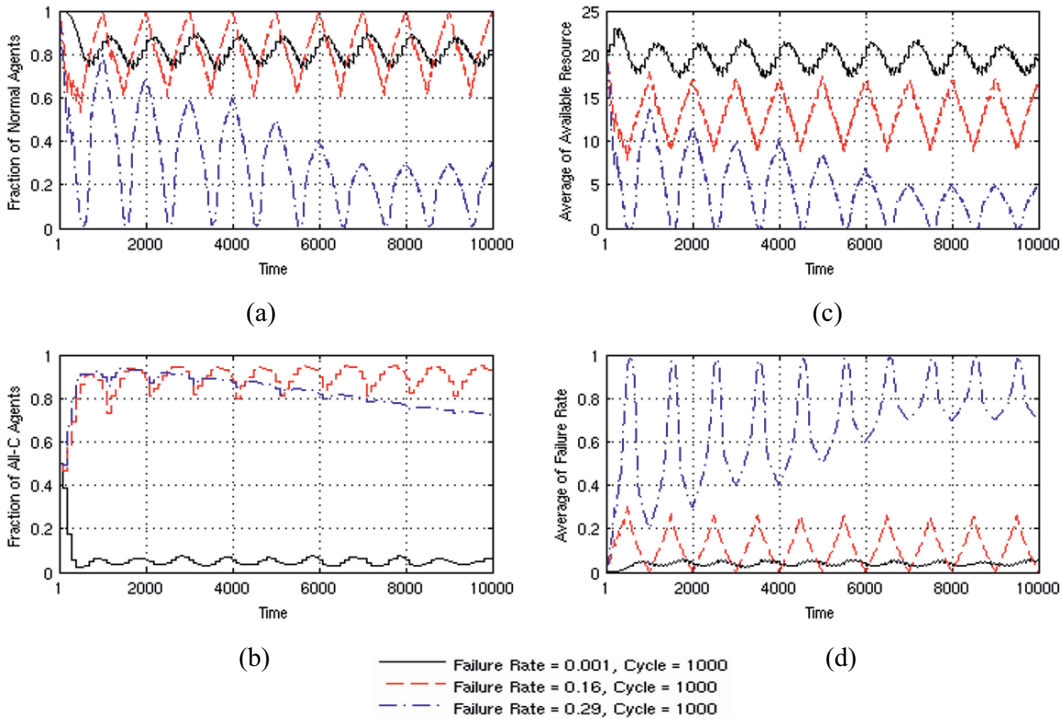


Fig.3 The performance development when the failure rate changes as a triangle wave. (a) fraction of normal agents (b) fraction of All-C agents (c) the average of available resource per agent (c) the average of failure rate in the network Parameters are as in TABLE 1 and initial configuration with the strategy each agents and 100 failure agents randomly chosen. The failure rate is between 0.001 to 0.90

D. Performance when the failure rate is kept constant

. The performance in a static environment when the failure rate is kept constant has been first studied. Figure 2 plots performance versus the failure rate. A threshold of the amplitude can be observed about the value 0.16. When the failure rate exceeded the threshold, both the fraction of normal agents (a) and the average of available resources decreased to 0 drastically. In contrast, the averaged failure rate (d) as well as the fraction of All-C (cooperating) agents increased. In addition, the average of available resource (c) also decreased to 0 slowly as the failure rate increases.

Although the fraction of All-C agents gradually increased below the threshold, causing the repair frequency increases, the fraction of All-C agents goes to a steady value about 0.5 when the threshold is exceeded.

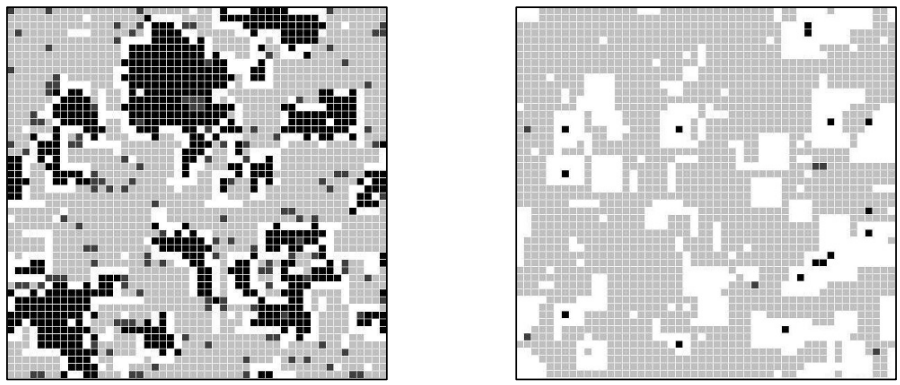
In sum, the performance measured by the fraction of normal agents and by the averaged available resource remained in a reasonable level when the failure rate is less the threshold. The agents adapted their actions to the environment. However, they ceased to make abnormal agents repair when the failure rate exceeded the threshold. In that situation, agents found that it would be useless to repair neighbour agents, and about a half of agents remained silent.

E. Performance when the failure rate changes dynamically

Since strategic repair allows each agent to take its own action, it generally permits a better performance than the case when the repair rate is fixed in time and uniform among agents, particularly in a heterogeneous environment where each agent fails with different failure rate. The strategic repair is also expected to perform well in a dynamic environment with a time-varying failure rate.

A time-varying function is used to specify the failure rate at any real valued time. We conducted computer simulations to investigate the performance in a dynamic environment when the failure rate oscillates with certain amplitude and cycle First, we consider the case when the failure rate oscillates with a triangle wave as follows,

$$f(t)=\begin{cases}\dfrac{2A}{T}t & \left(0\le t\le\dfrac{T}{2}\right)\\ -\dfrac{2A}{T}(t-T)+2A & \left(\dfrac{T}{2}<t\le T\right)\end{cases} \tag{1}$$



(a) Time step 2198 (b) Time step 3010

Fig.4 A snapshot of agent configuration when the simulation is carried out with the same condition as that in Fig. 3. Failure rate is 0.1 and the cycle is 1000. Light gray is Normal Cooperator, dark gray is Abnormal Cooperator, white is Normal Defector and black is Abnormal Defector.

where A is the amplitude and T is the cycle of the function. When a damage rate δ is added to a failure rate, the failure rate $\lambda(t)$ can be expressed with the number n counting the number of repair by abnormal agents as follows.

$$\lambda(t) = f(t) + n\delta \quad (2)$$

Figure3 plots performance evolution when time develops with the triangle wave. When the failure rate is 0,001 and 0.16, the performance behaves similarly to the change of failure rate. However, the performance measured by the fraction of normal agents and by the averaged available resource decreases when the failure rate is high, as plotted in (a) and (b) when the rate is 0.29. The agents adapted to the environment when the failure rate is less than the threshold.

Fig.4 shows a snapshot of an agent configuration when the simulation is carried out with same condition as that with Fig.3 except the failure rate 0.1, and the cycle 1000. The time step 2198 is near peak of the failure rate function (1). In Fig.3 (a), there are some black clusters of abnormal defectors which are surrounded by white normal defectors. However, the clusters of normal cooperator (light gray) prevent the abnormal agent clusters from expanding. If the agents of abnormal defector fail in repair, the agents of normal cooperator can repair them. The time step 3010 is the initial of the failure rate function (1). Fig.3 (b) shows that there are some abnormal defectors. However, since the clusters of normal defectors are small. thus, preventing the abnormal clusters from expanding.

F. Performance versus the failure rate

Figure 5 plots the performance with different the failure rate (legend). Because the failure rate has a threshold around the value 0.16, the performance measured by the fraction of normal agents (a) and by the averaged available resource (c) indicated a threshold near the value.

The fraction of normal agents (a) is kept constant 0.8 between the values 0.001 and 0.1. Further, the average of available resource (c) decreases relatively slow between these values. The fraction of All-C agents increases (b) in order to maintain the network. The gradual increase in the number of cooperating (hence repairing) agents (b) indicates that the agents adapt to the environment well. Due to the gradual increase of repairing agents, the averaged available resource accordingly decreases (c).

The fraction of the normal agent decreases drastically (a) when the failure rate is greater than 0.2. The average of failure rate in the network also increases (d) due to the addition of damage rate by repairing done by the abnormal agents. Accordingly, available resource decreased to 0 (c). However the performance changes differently depending on cycle when the failure rate is greater than the threshold 0.2. Further, the network can perform relatively better when the cycle is high even when the threshold is exceeded.

G. Performance versus the cycle

Figure 5 plots the performance varying the cycle. When the failure rate is greater than the threshold, the performance of the network has been affected by the cycle. While on the other hand when the failure rate is less than the threshold, the performance does not depend on the cycle. When the failure rate exceeded the threshold, the fraction of normal agent decreases to 0. However, as the cycle becomes greater than 1000, the fraction of normal agent improved slowly.

When the cycle is between 10 and 100, the failure rate increases drastically, however, the value decreased after the cycle is greater than 1000. High failure rate can be compensated by a high frequency.

It can be also observed that performance measured by the fraction of All-C agent (b) and by the averaged available resource remain independent from the cycle (b), when the

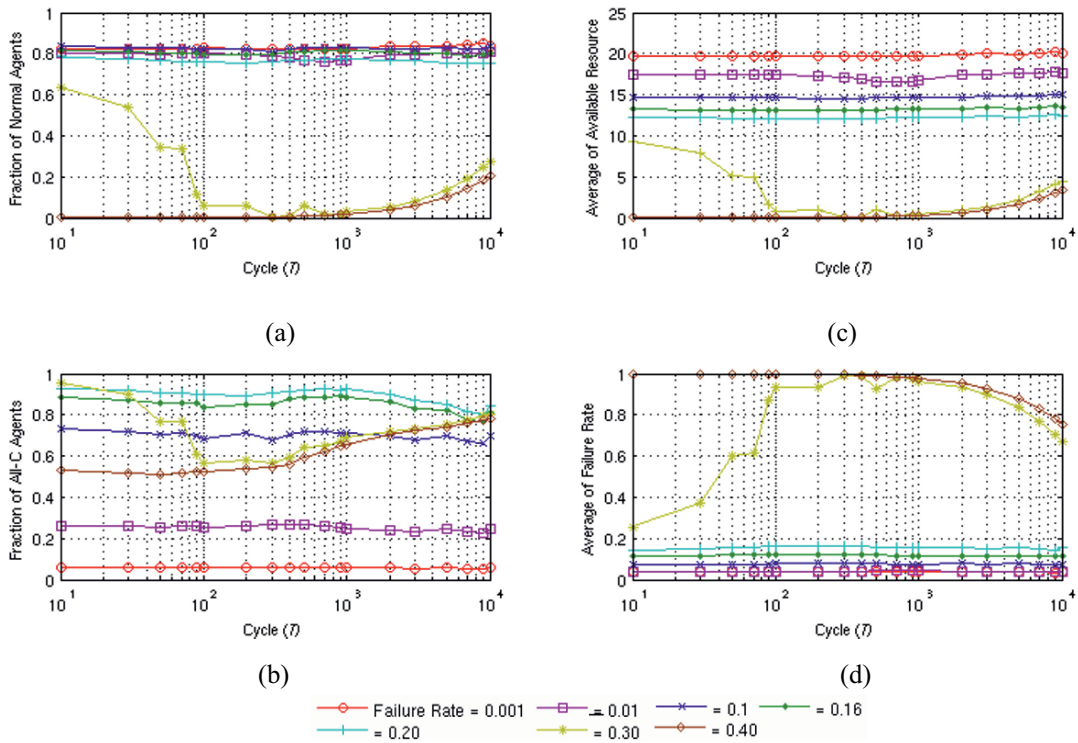


Fig.5 The performance versus cycle when the failure rate changes as the triangle wave. (a) fraction of normal agents (b) fraction of All-C agents (c) the average of available resource per agent (c) the average of failure rate in the network Parameters are as in TABLE 1 and initial configuration with the strategy each agents and 100 failure agents randomly chosen. The failure rate is between 0.001 to 0.90

failure rate is less than the threshold. However, the performance degraded as the cycle increases.

III. Related Works and Discussions

There have been many works for self-organizing and self-repairing of information systems. To mention only two in the spectrum of researches: theoretical one on one extreme and practical one on the other end of extreme. On the one hand, theoretical studies using cellular automata for self-organization of systems have been made [7]. The work is motivated by a problem of statistical physics, but it states the result for attaining reliable systems using probabilistic cellular automata. On the other hand, studies and development for Recovery-Oriented Computing have been made [8] where practical problem of rasing availability of computers is focused, involving hardware platform. It often happens that it would cost very high to attain high availability and reliability in information systems [9]. The study of recovery-oriented computing as well as self-repairing and self-organization by autonomous agents are motivated in part by this high cost problem.

In our previous study [3], computer simulation revealed that the strategic repair allows adaptation to the heterogeneous environment where each agent has distinct failure rate. The computer simulations of this paper revealed, with the strategic repair, the self-repairing network has been shown to adapt to a dynamic environment where the failure rate dynamically changes. The network adapted to the environment showed a reasonable performance when the failure rate is less than the threshold.

Dynamic failure rate is one property to be handled carefully to attain high reliability and availability in information systems. Although estimating and modelling the failure rate for mechanical systems seems relatively established (e.g. the one using Weibull distribution) with the study of mechanical wear-out and accumulation of fatigues, modelling the dynamics failure rate for information systems is difficult problem. This is because design failures are easy to get into information systems, and difficult to be found. Importantly,

malicious tempering can be done in many ways. Theoretical studies for modelling and estimating dynamic failure rate for information systems have been involved in the studies of network queue management, load-balancing,

IV. CONCLUSION

We considered the self-repairing network in a dynamic environment where the failure rate is a time-varying function; a triangle wave function in our computer simulations. As a result, the network performed well when measure by the fraction of normal agents and average of available resource. The simulation also demonstrated that the strategic repair allows adaptation to the dynamic environment.

ACKNOWLEDGMENT S

This work was supported by The Global COE Program "Frontiers of Intelligent Sensing", from the ministry of Education, Culture, Sports, Science and Technology. This work was also supported in part by Grants-in-Aid from Toyohashi University of Technology.

References

[1] R.K. Sahoo, A. Sivasubaramanian, M.S. Squillante, and Y, Zhang: Failure data analysis of a large scale heterogeneous server environment, Proc. Intl. Conf. Dependable Systems and Networks, pp.772-781, 2002

[2] D. Tang, R.K, Iyer, and S.S. Subramani: Failure analysis and modelling of VAXcluster system, Proc. Intl. Symp. Fault-tolerant Computing, pp.244-251, 1990.

[3] M. Oohashi, and Y. Ishida, A Game Theoretic Approach to Regulating Mutual Repairing in a Self-Repairing Network,, in *Innovative Algorithms and Techniques in Automation, Industrial Electronics and Telecommunications*, Springer, Netherlands, pp. 281-286, 2007.

[4] M. Nowak, and RM. May, "Evolutionary games and spatial chaos." Nature, 359, 29, 826-9, 1992

[5] P. Grim, "The greater generosity of the spatialized prisoner's dilemma," J. theor. Biol., Vol. 173, pp. 353-359, 1995.

[6] Y. Ishida, and T. Mori, "Spatial Strategies on a Generalized Spatial Prisoner's Dilemma, " J. of Artificial Life and Robotics, Volume 9, Number 3, pp. 139 - 143, 2005.

[7] P. Gacs, "Reliable Cellular Automata with Self-Organization," J. Stat. Phys. 103, pp. 45-267, 2001

[8] D. Oppenheimer, A. Brown, J. Beck, D. Hettena, J. Kuroda, N. Treuhaft, D. A. Patterson, K. Yelick, "ROC-1: hardware support for recovery-oriented computing," IEEE Trans. on Computers, Volume: 51, Issue: 2, pp. 100-107, 2002.

[9] W. Sawyer, "Case studies from hp's 5nines program: The cost of moving from 99% to 99.999% availability," In Second Workshop of the High-Dependability Computing Consortium (HDCC), Santa Cruz, CA 2001.

Information Sharing Between CSIRT and IDS

Zair Abdelouahab & Fernando A. Pestana Júnior
Federal University of Maranhão, CCET/DEEE
Av. Dos portugueses, Campus do Bacanga, São Luis – MA 65080-040
(zair, pestana)@dee.ufma.br

***Abstract*- This article proposes a message format for sharing information between Computer Security Incident Response Team (CSIRT) and Intrusion Detection Systems (IDS), aiming the achievement of an automatic update of the response action data base in IDSs, based on restrictive short-term measures suggested in security alerts issued by CSIRTs. This model is based on Web services and Extensible Markup Language (XML) technologies. It is also presented a data format to these security alerts as an extension of the Common Alerting Protocol..**

I. INTRODUCTION

Intrusion Detection Systems (IDS) are systems composed of software or hardware which automate the monitoring process of events that occur in a computing system or network, with the aim of identifying signs of security problems [1].

Mechanisms which allow exchanging information between an IDS and a Computer Security Incident Response Team (CSIRT) are necessary. Thus, a better understanding of events surrounding the protected domain allowing the prevention and response measures are taken.

In this work, it is proposed a model for sharing information between CSIRT and IDS, aiming to make it possible for an IDS to update automatically its response action data base, based on short-term restrictive measures suggested in security alerts issued by CSIRT, using Web services [10] and XML [9] technologies.

II. PROPOSAL OF A MODEL FOR SHARING INFORMATION BETWEEN CSIRT AND IDS

In this section, a model for sharing information between CSIRT and IDS is proposed, with the objective of enabling an IDS to automatically update its set of response actions, based on short-term restrictive measures suggested in security alerts released by CSIRT using Web services and XML technologies.

A. Response Actions Data Base (RADB)

The RADB contains information about response actions which are taken according to the detected attack. The scheme of RADB is a representation of the standard format, proposed to an alerts released by CSIRT which will be described in this section.

A response may be defined as a plan of action which the system has when an intrusion is detected [2]; this plan of actions will constitute the RADB. The list of main response activities are [7]: generate reports, alarms, cancel jobs and cancel user's session, investigate suspects, blocking IP addresses, disconnect the host, use a additional intrusion detection tool, disable ports and affected services, investigate the connection, create backup and use temporary file for protection.

B. Updating RADB based on Alerts released by CSIRT

According to Figure 1, a Central Security Agency is responsible for collecting alerts released by several CSIRTs, which have information about new vulnerabilities detected and they are made available in some standardized format. This information sharing by CSIRT, apart from the current pages in HyperText Markup Language (HTML), it may be done by means of Web services and XML which can be consulted constantly by the Central Security Agency searching for new alerts.

As long as the Central Security Agency receives a new alert released by a CSIRT, it will be processed and its information is stored in the RADB.

In figure 1, there are some CSIRTs which emit security alerts and can be used as information source to update the RADB.

When the alerts are published in HTML, the security administrator must read all released alerts and verify which published vulnerabilities are applicable in his environment and take the suggested measures in order to stop an intruder's access with the purpose of compromise the system, limit the extent of an intrusion, and prevent an intruder from causing further damage [3].

These restrictive short-term measures are exactly the ones which can feed the RADB, keeping the IDS response mechanism updated and able to respond to security incidents which involve a new vulnerability described in the alert released by CSIRT. In order to make it possible, it is necessary that the attack detection mechanism is also updated. Updating the attack detection mechanism is out of the scope of this work.

In order to accomplish the information sharing between CSIRT and IDS, it is necessary to have an agreement between them about the set of elements which they will use and know what these elements mean, and what Web services methods are going to use, what is the function of these methods and in which order they are called when more than one method is necessary.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 561–565.

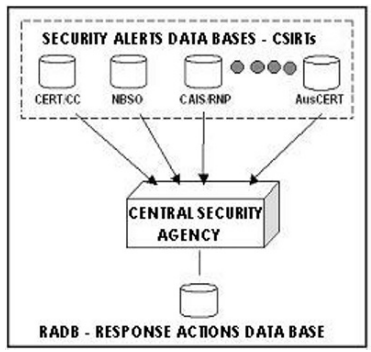


Fig. 1. Model for update automatically the response action data base

Apart from these circumstances it is necessary a standardization of CSIRT security alerts and this standard must allow automatic processing. With the standardization, an IDS does not have to understand and decode several formats. We propose an alert codification format as an extension of CAP

The structure of this alert message concerning security problems on the Internet is based on the structure of the CAP alert message and consists of one segment <alert> which can contain one or more segments <restriction> and <info>. The segments <info> contains one segment of <instruction> which in turn may contains one or more segments of <applyPatch>, <removePatch>, <blockAccess>, <disableFeature>, <modifyFile> and <PermissionChange>. The structure is shown in figure 2.

The following extensions are inserted in the structure of the CAP alert message:

- The required element “scope” of the CAP message must have its value determined as “Restricted”, pointing out that a certain alert is destined to users of specific systems.
- The optional element “restriction” is redesigned as a required segment which will point out which systems an alert refers to. Figure 3 shows the alteration.
- The optional element “instruction” contained in the segment “info” is redesigned as a required segment and contains suggestions of action to be taken. Figure 4 shows the structure of those elements.

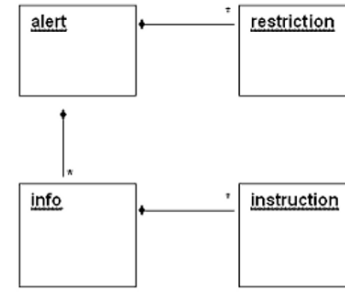


Fig. 2. Structure of the alert message for security problems on the Internet.

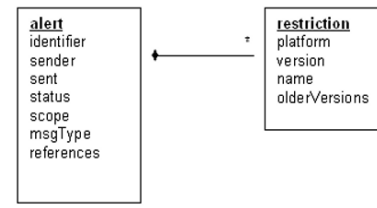


Fig. 3. Structure of segments “alert” and “restriction”.

.

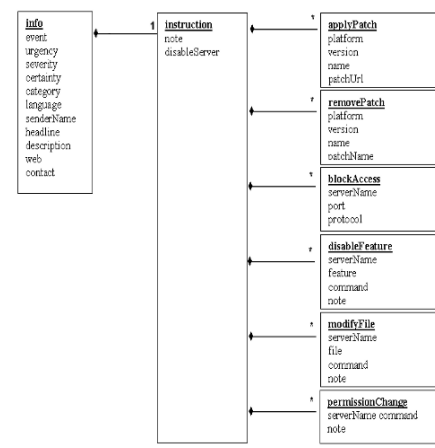


Fig. 4. Structure of elements “info”, “instruction” and extensions.

C. Data Dictionary

TABLE I
"RESTRICTION" ELEMENT AND SUB-ELEMENTS

Element Name	Definition and (optionals)	Notes or Value Domain
restriction	The container for all component parts of the restriction sub-element of the alert message. (required)	Multiple occurrences are permitted within a single <alert>. Identifies vulnerable systems.
platform	The code denoting the platform name of the vulnerable system. (optional)	Code Values: "Sparc" – Sun Sparc Platform. "x86" – Intel Platform. "Windows" – Microsoft Windows. "Solaris" – Sun Solaris. "Linux" – Linux Platform.
version	System version identification. (optional)	A number or string uniquely identifying the system version, assigned by the vendor. No spaces or restricted characters (< and &).
name	The identifier of the system name. (required)	A string uniquely identifying the system name, assigned by the vendor. No spaces or restricted characters (< and &).
Older Versions	The code denoting if an alert affects older versions of a specific system. (optional)	Code Values: "Y" - An alert affects all older versions of a specific system. "N" - An alert does not affect all older versions of a specific system.

TABLE II
"INSTRUCTION" ELEMENT AND SUB-ELEMENTS

Element Name	Definition and (optionals)	Notes or Value Domain
instruction	The container for all component parts of the instruction sub-element of the info sub-element of the alert message. (required)	Single occurrence permitted within a single <info> block. In addition to the specified sub-elements, may contain one or more <applyPatch>, <removePatch>, <blockAccess>, <disableFeature>, <modifyFile>, <permissionChange> blocks.
note	The text describing the recommended action to be taken by recipients of the alert message. (required)	
Disable Server	The group listing of server names to be disabled.	Multiple space-delimited server names may be included. Server names including white space must be enclosed in double-quotes. A string uniquely identifying the server name, assigned by the vendor.

TABLE III
"APPLYPATCH" ELEMENT AND SUB-ELEMENTS

Element Name	Definition and (optionals)	Notes or Value Domain
applyPatch	The container for all component parts of the applyPatch sub-element of the instruction sub-element of the info sub-element of the alert message. (optional)	Refers to a patch file. Multiple occurrences are permitted within a single <instruction> block.
platform	The code denoting the platform name of the vulnerable system. (optional)	Code Values: "Sparc" – Sun Sparc Platform. "x86" – Intel Platform. "Windows" – Microsoft Windows. "Solaris" – Sun Solaris. "Linux" – Linux Platform.
version	System version identification. (optional)	A number or string uniquely identifying the system version, assigned by the vendor. No spaces or restricted characters (< and &).
name	The identifier of the system name. (required)	A string uniquely identifying the system name, assigned by the vendor. No spaces or restricted characters (< and &).
patchUri	The identifier of the hyperlink for the patch file. (required)	A full absolute URI, typically a Uniform Resource Locator that can be used to retrieve the patch file over the Internet.

TABLE IV
"REMOVEPATCH" ELEMENT AND SUB-ELEMENTS

Element Name	Definition and (optionals)	Notes or Value Domain
Remove Patch	The container for all component parts of the removePatch sub-element of the instruction sub-element of the info sub-element of the alert message. (optional)	Refers to an older patch that must be removed before install a new patch. Multiple occurrences are permitted within a single <instruction> block.
platform	The code denoting the platform name of the vulnerable system. (optional)	Code Values: "Sparc" – Sun Sparc Platform. "x86" – Intel Platform. "Windows" – Microsoft Windows. "Solaris" – Sun Solaris. "Linux" – Linux Platform.
version	System version identification. (optional)	A number or string uniquely identifying the system version, assigned by the vendor. No spaces or restricted characters (< and &).
name	The identifier of the system name. (required)	A string uniquely identifying the system name, assigned by the vendor. No spaces or restricted characters (< and &).
patchName	The identifier of the patch name. (required)	A string uniquely identifying the patch name, assigned by the vendor. No spaces or restricted characters (< and &).

TABLE V
"BLOCKACCESS" ELEMENT AND SUB-ELEMENTS

Element Name	Definition and (optionals)	Notes or Value Domain
blockAccess	The container for all component parts of the blockAccess sub-element of the instruction sub-element of the info sub-element of the alert message. (optional)	Refers to ports and protocols that must be blocked. Multiple occurrences are permitted within a single <instruction> block.
serverName	The identifier of the server where ports or protocols will be blocked. (required)	A string uniquely identifying the server name, assigned by the vendor. No spaces or restricted characters (< and &).
port	The integer from 0 through 65535 indicating the port number to be blocked. (required)	Port numbers assigned by IANA (Internet Assigned Numbers Authority).
protocol	The code denoting the protocol name. (required)	Code Values: "Tcp" – Transmission Control Protocol. "Udp" – User Datagram Protocol.

TABLE VI
"DISABLESERVICE" ELEMENT AND SUB-ELEMENTS

Element Name	Definition and (optionals)	Notes or Value Domain
disableService	The container for all component parts of the disableService sub-element of the instruction sub-element of the info sub-element of the alert message. (optional)	Identifies a service or feature of this service that must be disabled. Multiple occurrences are permitted within a single <instruction> block.
serverName	The identifier of the server where a service will be disabled. (required)	A string uniquely identifying the server name, assigned by the vendor. No spaces or restricted characters (< and &).
service	The identifier of the service that will be disabled. (required)	A string uniquely identifying the service or a feature of this service, assigned by the vendor. No spaces or restricted characters (< and &).
command	The text describing a command and its options to disable a service or a feature of this service.	
note	The text describing the service or a feature of this service to be disabled and probably impacts of these actions.	

TABLE VII
"MODIFYFILE" ELEMENT AND SUB-ELEMENTS

Element Name	Definition and (optionals)	Notes or Value Domain
modifyFile	The container for all component parts of the modifyFile sub-element of the instruction sub-element of the info sub-element of the alert message. (optional)	Refers to a file name with its full path. Typically a configuration file. Multiple occurrences are permitted within a single <instruction> block.
serverName	The identifier of the server where a file or files will be modified. (required)	A string uniquely identifying the server name, assigned by the vendor. No spaces or restricted characters (< and &).
file	File name with its full path. (required)	
command	Text describing the necessary command to modify a file. (optional)	Multiple occurrences are permitted within a single <modifyFile> block.
note	Text describing the command and probably impacts of its actions. (optional)	

TABLE VIII
"PERMISSIONCHANGE" ELEMENT AND SUB-ELEMENTS

Element Name	Definition and (optionals)	Notes or Value Domain
permissionChange	The container for all component parts of the permissionChange sub-element of the instruction sub-element of the info sub-element of the alert message. (optional)	Used to modify directory or file access permission. Multiple occurrences are permitted within a single <instruction> block.
serverName	The identifier of the server where directory or file access permission will be modified. (required)	A string uniquely identifying the server name, assigned by the vendor. No spaces or restricted characters (< and &).
command	Text describing the necessary command to modify a user permission to access directories or files. (required)	Multiple occurrences are permitted within a single <permissionChange> block.
note	Text describing the command and probably impacts of its actions. (optional)	

III. NIDIA System

NIDIA is a proposal of a multiagent [8] IDS, able to analyze data from hosts logs and network traffic packets to generate an attack suspicion level of the protected network.

The NIDIA architecture is inspired in the CIDF logical model, and has agents acting as event generators (sensor agents), data analysis mechanisms (system monitoring and security evaluation agents) and response module (system controller agent). In addition, there are agents accountable for system integrity (system integrity agent) and system update (system update agent).

There are also databases to store its security policy strategy (STDB), the response actions to be taken in case of suspicious activity (RADB), intrusion and intruder patterns (IIDB) and a database to store intrusion data (DFDB).

Aiming to keep the response actions database updated, the System Updating Agent is implemented according to the multiagent architecture proposed in [5] for NIDIA System. The agent works as a Central Security Agency, as described in section 4.2.

In order to do some tests, a prototype representing a CSIRT is implemented, which makes several security alerts available through a web service. A database is developed for this prototype and it is constantly updated with CERT®/CC vulnerability notes released in 2004.

With this implementation it is possible to maintain updated the NIDIA response actions data base. The System Updating Agent constantly sends requests to the prototype of CSRIT which processes them and in case there is a new alert, it sends to NIDIA the alerts in XML format.

When an intrusion is detected by NIDIA, controller agent of NIDIA responds according to the data updated and stored in RADB.

IV. Conclusion

With daily discovery of new threats and vulnerabilities in computing systems, updating IDS becomes a constant worry.

This work proposes a means of how to keep the Response Actions Data Base (RADB) updated through information share between IDSs and CSIRTs.

However, in order to share information it is necessary a standard format of the alert, which is proposed as an extension of Common Alerting Protocol (CAP).

Using a prototype, based on society of intelligent agents, Web services and Extensible Markup Language – XML technologies, it is possible to establish information sharing and keep the RADB update, which proves the viability of the model herein presented.

Acknowledgment

Financial support of FAPEMA is gratefully acknowledged.

References

[1] R. Bace, P. Mell. “Intrusion Detection Systems”. *NIST Special Publication*, SP800-31, (1999).

[2] R. Bace. *Intrusion Detection.* Macmillan Technical Publishing, (1999).

[3] Carnegie Mellon University / Software Engineering Institute. *Responding to Intrusions*. CMU/SEI-SIM-006, February (1999).

[4] CIDF Working Group. *The Common Intrusion Detection Framework Architecture*. Draft CIDF (1998).

[5] C.Lima et al, “The NIDIA Project Network Intrusion Detection System based on Intelligent Agents”, *Proceedings of Tenth Latin-Ibero-American Congress on Operations Research and Systems*, Mexico , pp. 212-217, 2000.

[6] Organization for the Advancement of Structured Information Standards. *Common Alerting Protocol,* v. 1.0. OASIS Standard 200402 (2004).

[7] G. Santos, E,Nascimento. “An Automated Response Approach for Intrusion Detection Security Enhancement” *Proceedings of VII International Conference on Software Engineering and Applications*. Marina Del Rey, California, USA 2003.

[8] G. Weiss, *Multiagent Systems - A Modern Approach to Distributed Artificial Intelligence,* The MIT Press - Cambridge, Massachusetts, London, England (1999).

[9] World Wide Web Consortium. *Extensible Markup Language (XML) 1.1*. April (2004).

[10] World Wide Web Consortium. *Web Services Architecture.* February (2004).

Cellular Automata Used for Congestion Control in Wireless LANs

Zornitza Genova Prodanoff

School of Computing
University of North Florida
zprodano@unf.edu

***Abstract*—This study describes a novel technique of using cellular automata for congestion control in wireless LANs (WLANs). Our approach relies on cooperation between central nodes in WLANs, covering the same geographical area. Such geographical areas, even in the scope of a single building, may be covered by several available wireless networks, supporting same or different WLAN standards over the same space. We propose a technique for congestion avoidance in such heterogeneous environments, based on the use of cellular automata (CA). When each mobile device attempts to join a wireless network, a decision is made by the central node, whether to accept the association attempt, or reject it, in accordance with the outcome of a CA execution. Our technique attempts to reach a social goal for a large number of WLANs of optimizing the use of central node resources. This technique is not applicable in ad hoc networks due to the inherent decentralization of resources.**

Index Terms — WLAN, congestion control, cellular automata, IEEE 802.11.

I. Introduction

Wireless devices, supporting the IEEE 802.11 suit of standards (WiFi) are widely available today on the consumer market – personal digital organizers, laptop computers, tablet PCs, smartphones, etc. This trend leads to the development of supporting networks, worldwide. There are projects initiated to install Wireless LANs in big clusters and cover entire cities. Examples in the US are projects in the following cities: Grand Haven, Michigan, Grand Rapids, Michigan, Boston Main Street WiFi initiative, Massachusetts, Philadelphia, Pennsylvania, Tempe, Arizona, Dayton, Ohio, the initiated by GOOGLE Free City Wide WiFi in San Francisco, California. Other projects are initiated outside the US, such as in Amsterdam, Netherlands and Taipei, Taiwan. Over 300 cities have already had their city-wide WiFi LAN and the number keeps growing [10].

In addition, 802.16 based wireless MAN projects are initiated for city wide wireless access coverage, such as the related city of Tokyo project [5]. Laptops and other mobile devices with integrated IEEE 802.16 (WiMAX) interfaces are expected to be widely available on the consumer market in the next few years. Dan Coombes, Motorola Networks senior vice-president, predicts that 10 million WiMAX laptops will ship in 2008. [7]

In parallel with IEEE 802.11 and IEEE 802.16 networks, there are 2G (e.g. EDGE) and 3G GSM and CDMA mobile phone data network protocols that consumers can use to access the Internet from smartphones. They are available in many non-urban and remote areas. Cell phone providers already allow users to choose between their own mobile phone network and an IEEE 802.11 hotspot to access the Internet, or to make phone calls.

Coverage by IEEE 802.16 based networks complements very well the widely available IEEE 802.11 based LANs. Metropolitan Area Networks, such as IEEE 802.16 are easier to build and maintain in areas, located outside main urban centers. Stationary mobile users, or users that are not in motion for the duration of the application session, have a variety of choices to decide, which of a number of available LANs or MANs to join.

For cell phone users who are mobile and roaming, e.g. traveling in a car, while accessing the Web, cell phone networks solve the handoff problem. There are existing IEEE 802.11 based techniques that provide continuous connectivity to a roaming client [9]. However, there are no widely deployed and efficient, transparent handoff solutions for wireless 802.11 networks [1], [4]. Existing networks require specially modified client software to be able to transfer TCP connections from one access point to the next.

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 566–570.

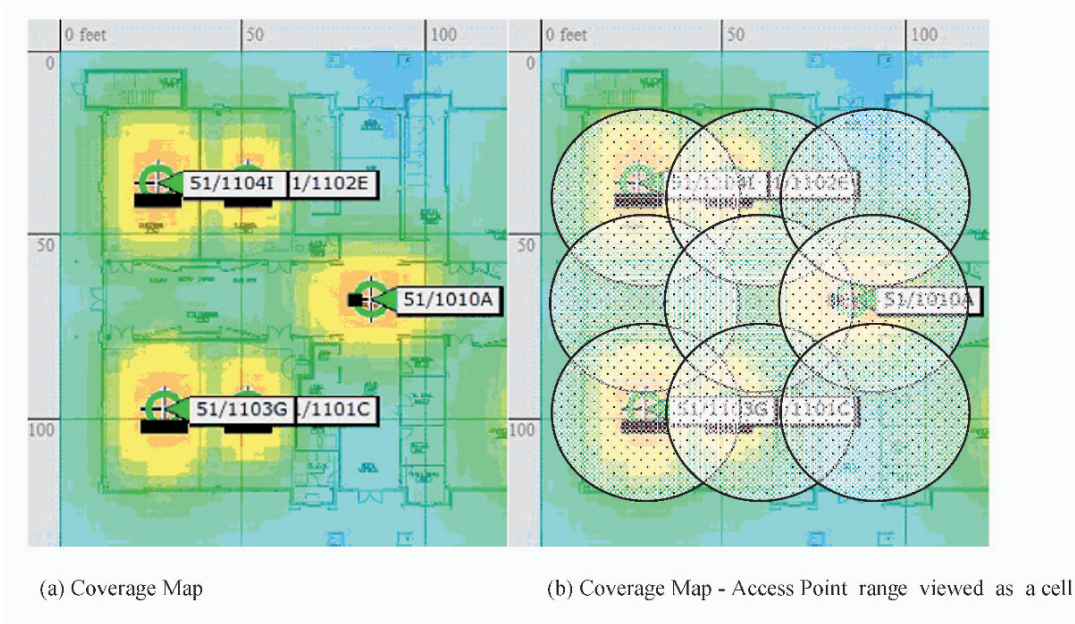


Figure 1. IEEE 802.11b/g Coverage Map

In this study we investigate the problem of cooperation between central nodes in wireless local area networks (WLANs), covering the same geographical area. As shown in Figure 1, the same geographical area, even in the scope of a single building, may be covered by several available WLANs. Figure 1 depicts a single building layout with five available IEEE 802.11b/g access points, covering a joint area of over 30,000 square feet. We propose a novel technique for congestion avoidance in such environments, based on cellular automata. When each mobile device attempts to join a WLAN, a decision will be made by the central node, whether to accept the join attempt, or reject it. A cellular automaton is then executed to help in the decision making.

This paper is organized as follows. Section II provides an overview of the problem of congestion in some existing wireless network protocols. Section III presents the use of cellular automata for congestion control in large scale heterogeneous wireless networks environments, where several wireless standards and multitude of their instances are employed over the same geographical area. Section IV provides a summary of this study and outlines future work directions.

II. Congestion In Wireless Networks

Traffic congestion in a large system of several nearby positioned WLANs is an interesting problem, since resources in wireless networks are limited, leading to degraded performance of existing traffic congestion techniques for wired networks.

The performance of the various protocol layers has been shown before to be affected by congestion. In wireless networks, both the lossy nature of the channel and congestion contribute to degraded performance, however there are existing algorithms based techniques to help distinguish congestion from wireless network losses [3]. Reactive optimization techniques to improve the performance of IEEE 802.11 protocol by adjusting frame sizes have been introduced, and studied before [6], [8]. The effect of congestion on frames transmission time, when frames are sent over channels supporting different data rates have been theoretically analyzed before [2]. The results in [2] are also confirmed in [11]. Metrics to estimate congestion levels are discussed and proposed in [11]. Link reliability and channel utilization (measured as busy-time) along with network throughput can be used to detect congestion.

We propose a proactive traffic congestion alleviation technique for infrastructure based (as opposed to ad hoc) wireless networks that relies on cooperation between central nodes (access points) and estimation of current congestion levels. Our technique can also be used when bit error rates are high, since it is threshold based and the threshold value is estimated by tracking channel utilization and network throughput, as well as bit error rates.

III. Cellular Automata Used for Congestion

CONTROL IN LARGE SCALE HETEROGENEOUS WIRELESS NETWORKS

Let A be a finite set, called the set of states. We denote with d, $d > 0$ the *dimension* of the cellular automaton. The finite set $N \subseteq \mathrm{Z}^d$, where Z^d is a $d-$dimensional integer lattice, is used to specify the local neighborhood of the cellular automaton. For $v \in \mathrm{Z}^d$ we will denote $v+N$ the set $v+N=\{v+j \mid j \in N\}$.

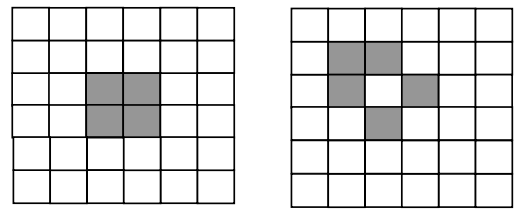
If $d=2$ and $N=\{-r,...,-1,0,1,...,r\}$ then for $v \in \mathrm{Z}$ we have that $v+N$ is the set of all elements in Z^2 which are on distance at most r from v (in this case the distance of (i,j) and (i',j') is the $\max\{|i-i'|,|j-j'|\}$.

In general, for a dimension d if $N=\{(i_1,...,i_d) \mid 0 \le i_j \le r, j=1,...,d\}$ and $v \in Z^d$ we have that $v+N$ is the set of elements in Z^d which are on distance at most r from v. The neighborhood N is called $d-$dimensional neighborhood of radius r. It is a square box with side $2r+1$ centered at the origin. In the case of $r=1$ and $d=2$, the $2-$dimensional neighborhood of radius 1 is called the Moore neighborhood. Or, a cell and its eight immediate neighbor cells form a 3 × 3 cell square, $N = \{\{-1,-1\},\{0,-1\},\{1,-1\},\{-1,0\},\{0,0\},\{+1,0\},\{-1,+1\},\{0,+1\},\{1,+1\}\}$. These relative cell positions can be viewed as the directions on the compass, respective of order: N = {NW, N, NE, W, Center, E, SW, S, SE}.

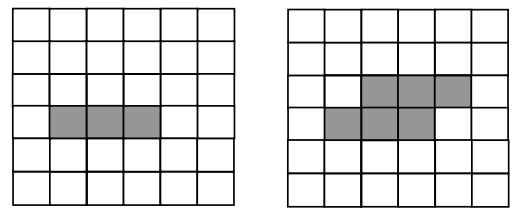
A global configuration is an assignment of a state to each element of Z^d, i.e. it is a function from Z^d to A and is denoted with one of the Greek letters. If M is a subset of Z^d and α is a global configuration, then α_M is the restriction of α on the index set M. The set of global configurations will be denoted with $\Gamma = \{\alpha : Z^d \to A\}$. A local function (also known as local dynamics function) is a function $l : A^N \to A\}$.

Definition: (*cellular automaton* [12]) A cellular automaton (CA) is a four-tuple $(A,d,N,l\}$, where A is a finite set of states and d is the dimension, N is a finite subset of Z^d called the neighborhood and $l : A^N \to A$ is the local function. The global function of the cellular automaton is $G_l = \Gamma \to \Gamma$ defined with $(G_l(\alpha))_v = l(\alpha_{v+N})$.

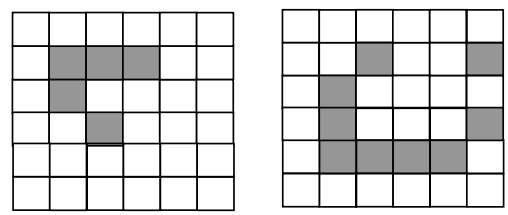
The Convay's Game of Life is a two-dimensional CA, where the neighborhood N is the Moore's neighborhood and the set of states is $A=\{0,1\}$, two states: *dead* or *alive* (graphically representing two colors used to color cells, gray and white in Figure 2.). The local function is defined with $l : A^N \to A$ with $l(w)=1$, if w contains three 1's, not in $\{0,0\}$ position (i.e. the cell becomes *alive* (state of 1) if there are three live cells surrounding it). Also, $l(w)=1$ if w contains two 1's not in position $\{0,0\}$ and $l(w)=0$ in the other cases, that is, the cell becomes *dead* (state of 0) if there is not enough life environment (only one or no cell in the neighborhood is alive) or if it is overcrowded (more than three live cells surrounding it).



a) Still "life form" examples



b) Oscillating "life form" examples



c) "Glider" – "life form" examples

Figure 2. Some Initial Game of Life Configuration Patterns

We can think of the automaton being *iterated*, that is, its local function computed over some initial configuration and then re-executed over the resulting configurations, until some halting condition is met. After the automaton is iterated and the halting condition is met, the last global configuration is examined, that is, it is computed whether all cells have died out completely, alive cells formed a stable population, oscillate, or the "live population" is growing forever. For some initial configurations (e.g., square 2×2 , see Figure 2), there is no change in consecutive iterations outcome, but other configurations die (e.g., a 3×3 square dies in three iterations). Some game of life configurations result in dying out patterns, where all live cells – both from the initial configuration and later "born" become dead after a finite number of iterations. Other starting patterns do not "die out", but become either stable figures or patterns that oscillate forever in a closed loop, see Figure 2. The reader is referred to [12] for more detailed presentation of CA.

Cellular automata have already been applied in real-time transportation traffic congestion simulations. Wireless network congestion control is aimed to solve a different problem. Unlike transportation traffic, wireless network traffic congestion is not estimated with a unit of space occupancy metric. It can be defined as load at the wireless routers or access points.

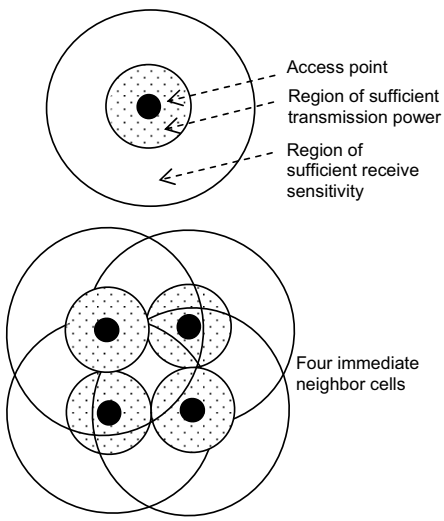


Figure 3. 802.11 WLAN cells model

We use the game of life (a two-dimensional cellular automaton) idea to model the network topology across multiple WLANS as Z^d , where $d = 2$. In such system each cell can be represented as a WLAN with certain transmission power and receive sensitivity. We will assume that two immediate neighbors are reachable form each other, that is, have sufficient levels of transmission power and receive sensitivity to support a full duplex communication link (See Figure 3). Figure 4 depicts how immediate neighbors are modeled to form the Moore's neighborhood.

If channel utilization and throughput in a WLAN exceed a threshold value (e.g., caused by too much traffic at an access point, or signal degradation (high bit error rates) caused by natural phenomena such as rain), we say that the cell becomes *alive*. If a new mobile node attempts to associate with this base station the cellular automaton is executed with some initial configuration, reflecting the state of all cells in the network topology. If such execution of the initial configuration results in a pattern that grows forever, the association request is declined.

```
I = number of rows;
J = number of columns;

alive          = TRUE;
dead           = FALSE;
cell[I,J]      = dead;
neighbor[I,J]  = dead;
update_matrix[I,J] = update from neighbor APs;

for (i=1 to I)
  for (j= 1  to J) // initialize configuration
  {
    if update_matrix(i,j)==alive;
    cell(i,j)=alive;
  }
while (halting condition not met)
{
for (i=1 to I) // begin iteration
  for (j= 1  to J) // for each cell
  {
    count = 0;
    position=0;
    for( position = 1 to 8) // for each neighbor
      if (neighbor[position] == alive)
        count++;

    if (cell(i,j) == alive)
      if !( (count == 2)||(count == 3) )
        cell(i,j) == dead;

    else // cell is dead
      if (count == 3)
        cell(i,j) == alive;

  } // end iteration

} // end execution
```

Figure 4. WLAN Congestion Avoidance Algorithm

The social goal is to reduce the total number of cells becoming alive (or base stations with performance problems) in the future. This is an attempt to predict (by executing the cellular automaton) if the traffic contributed from this new station will result in "global" network congestion, affecting more than one WLAN. It is hence necessary to share status information between neighbor access points. In this manner cellular automata can be used to improve aggregate network performance without having to introduce any changes to existing protocols at the wireless workstations.

The basic algorithm for implementing congestion control for each WLAN association request is presented in Figure 4. The halting condition in the pseudocode of Figure 3 is a union of the following cases:
- all cells die;
- system stable (oscillating pattern or still configuration);
- system unstable after some threshold reaching iteration step.

One of the interesting observations about the Game of Life CA is that implementing a fairly simple local function can be used to model quite complex global behavior. The time complexity of the above algorithm is $O(I \times J)$. Note that the complexity required to keep track and check if the halting condition is met is no worse than the above. Hence, we expect that a software implementation will be feasible in terms of added processing delay at the time of workstation association even in citywide WLAN cluster layouts.

IV. Summary and Future Work

This study describes a novel technique of using cellular automata for congestion control in WLANs. Our approach relies on cooperation between central nodes in WLANs, covering the same geographical area. Such geographical areas, even in the scope of a single building, may be covered by several available wireless networks, supporting same or different layers 1 and 2 protocols over the same space. We propose a technique for congestion avoidance in such heterogeneous environments, based on the use of cellular automata. When each mobile device attempts to join a wireless network, a decision is made by the central node, whether to accept the association attempt, or reject it. A cellular automaton is executed to support the decision making. Our technique attempts to reach a social goal for a large number of WLANs of optimizing the use of central node resources. This technique is not applicable in ad hoc networks due to the inherent decentralization of resources. The time complexity of the proposed basic algorithm for implementing congestion control in a group of WLANs depends on the number of cooperating WLANs.

References

[1] Y. Bejerano, I. Cidon, and J. Naor, "Efficient handoff rerouting algorithms: a competitive on-line algorithmic approach," ACM Transactions on Networking, vol. 10, no. 6, pp. 749-760, 2002

[2] G. R. Cantieni, Q. Ni, C. Barakat, and T. Turletti. Performance Analysis under Finite Load and Improvements for Multirate 802.11. Elsevier Computer Communications Journal, vol. 28, iss.10, pp.109-109, June 2005

[3] S. Cen, P. C. Cosman, and G. M. Voelker, "End-to-end Differentiation of Congestion and Wireless Losses." ACM Transactions on Networking, vol. 11, iss.5, pp.703-717, October 2003

[4] C F Chiasserini, "Handovers in Wireless ATM Networks: In-Band Signaling Protocols and Performance Analysis," IEEE Transactions on Wireless Communications, vol. 1, iss. 1, January 2002

[5] J. Libbenga, "Tokyo to enjoy city-wide WiMAX", Yozan MetroZone, March 2005. URL: http://www.theregister.co.uk/2005/03/18/tokyo-wimax/

[6] E. Modiano, "An Adaptive Algorithm for Optimizing the Packet Size Used in Wireless ARQ Protocols", Wireless Networks, vol. 5, iss.4, pp.279-286, July 1999

[7] P. Talacko, "City-wide hotspots increase options", June 21 2006. UTL: http://www.ft.com/cms

[8] M. Torrent-Morenoe, D. Jiang, and H. Hartenstein. Broadcast Reception Rates and Effects of Priority Access in 802.11 Based Vehicular Ad hoc Networks. In Proceedings of ACM VANET, Philadelphia, PA, October 2001

[9] H. Velayos and G. Karlsson, "Techniques to Reduce IEEE 802.11b MAC Layer Handover Time", KTH Technical Report, ISSN 1651-7717, pp.114-128, April 2003

[10] B. Wong, "Altai City-wide WiFi Cellular Network Solution", seminar, The Chinese University Of Hong Kong, Department of Information Engineering, February 2007. URL: http://www.ie.cuhk.edu.hk/fileadmin/seminar/2007/sem0607.pdf

[11] A. P. Jardosh, K. N. Ramachandran, K. C. Almeroth, and E. M. Belding-Royer, "Understanding Congestion in IEEE 802.11b Wireless Networks", proceedings of the Internet Measurement Conference, October 2005

[12] M. Garzon, "Models of Massive Parallelism: Analysis of Cellular Automata and Neural Networks", Springer-Verlag, London, UK, ISBN: 9783540561491, 1995

Performance Model of a Campus Wireless LAN

Seungnam Kang, Zornitza Prodanoff, and Pavan Potti
School of Computing
University of North Florida
{s.kang, potti.pavan, zprodano}@unf.edu

Abstract— **In this study we measured simulated end user HTTP response times in a 802.11g WLAN. Our results show that, ceteris paribus, modeling heavy tailed file sizes at an HTTP server, results in two times worse response times as compared to modeling Web pages of lognormal file sizes. Accesses to content containing large images and embedded documents appear to result in tolerable end user response times, as long as HTTP server directories contain a small number of such large files, that is, if file sizes are modeled by a heavy tailed Pareto distribution, where minimum file sizes are 100 Kbytes or less. When file size at the server become larger, with a minimum size of 1Mbytes, response times increase substantially and may quickly become intolerable by users. Our results confirm related findings in [3] and demonstrate that studies which do not consider these findings may present too optimistic results, misrepresenting real end user delays of Web accesses.**

Index Terms—Wireless LANs, congestion control, cellular automata.

I. Introduction

Applications that involve e-mail, Web browsing, and access to various server-based databases are commonly used today. The performance of these applications depends on the physical implementation of wireless LANs. The majority of campus WLANs are implemented to support the IEEE 802.11 WiFi family of protocols [5]. 802.11b provides communication at a theoretical maximum throughput of 11Mbps, while 802.11g and 802.11a at 54Mbps.

802.11b and 802.11g use the 2.4GHz frequency band for signaling. This frequency range is divided into 11 channels for use in the US. The effective width of each signal is about 11MHz either side of the nominal frequency. Channels 1, 6, and 11 have lower potential for interference coming from devices associated with other 802.11b/g access points, that is, these are "non-overlapping" channels. Communication between an 802.11 wireless card and an access point takes place over a common channel frequency, where the access point has a manually set frequency, and the wireless card of the user device "tunes into" the frequency of the available access points.

When more than one access point is in range, the wireless device may opt to associate with any available WLAN. If two users (devices) associated with two different access points within range are assigned to the same channel, the result is performance degradation. For this reason it is preferred that access points within range of each other use non-overlapping channels. With 802.11a this is not an issue because the 802.11a standard defines separate, non-overlapping channels. 802.11a is, however, less frequently chosen to be the technology of choice in a campus WLAN environment. This is due largely to its inherent downside of poor signal penetration properties trough objects such as walls that results in worse available range than comparable 802.11b/g implementations. Nevertheless 802.11a is still used when signal interference is a consideration, since it operates in the 5GHz spectrum and

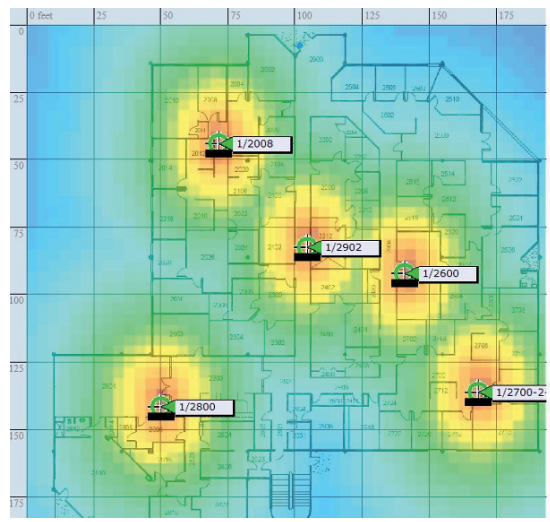


Figure 1. IEEE 802.11b/g Coverage Map

T. Sobh et al. (eds.), *Novel Algorithms and Techniques in Telecommunications, Automation and Industrial Electronics,* 571–576.

most wireless consumer devices (e.g. cell phones, PDAs, wireless headphones) are using the 2.4GHz frequencies. In this study we use 802.11b, since it is widely used and potentially, when a 802.11b device joins a 802.11b/g compatible LAN, the 802.11g maximum throughput drops automatically for all connected devices to the 802.11b supported maximum of 11Mbps.

This paper is organized as follows. Section II provides an overview of some prior work on evaluating wireless LANs. Sections III and IV present the OPNET simulation model we developed and evaluation results. Section V provides a summary of this study and outlines future work directions.

II. Measuring Wireless LAN Performance

A large number of studies have analyzed the performance of wireless networks. Most of them investigate transport and application layer performance through the analysis of traffic captured over an actual deployed WLAN implementation. Other studies rely on captured link-layer information to analyze the performance of small-scale campus wireless networks [13].

HTTP [10] performance over wireless physical layer implementations has been extensively studied in the past [1], [2], [4], [11], [12], [7].

Among these a few articles concentrate on campus-wide wireless networks (e.g. [6], [7]) and modeling Web browsing behavior. Our study has a similar goal – evaluating HTTP performance in a campus wireless network. We extend the idea in [7] to build a more realistic model based on heavy tailed Pareto distributed Web document sizes, where background traffic, originating from infrastructure based LANs, is generated and added to the model. We also study HTTP performance at extreme conditions, using "heavy Web browsing" patterns, defined in this study, based on prior simulation models from OPNET ITGuru 9.1 [8].

III. OPNET Simulation Model of a Deployed Campus LAN

A. Simulation Testbed

This study was completed on a Dell OPTIPLEX 2GX260 PC, running Windows XP Professional (2002 edition) and equipped with a 3.06GHZ Intel Pentium 4 processor and 1 GByte of RAM. The simulation software package used was OPNET ITGuru Academic Edition v. 9.1 (06/04/2007). The model library used was OPNET IT Guru Models Academic Edition 9.1.A PL1 (06/04/2007).

Among these

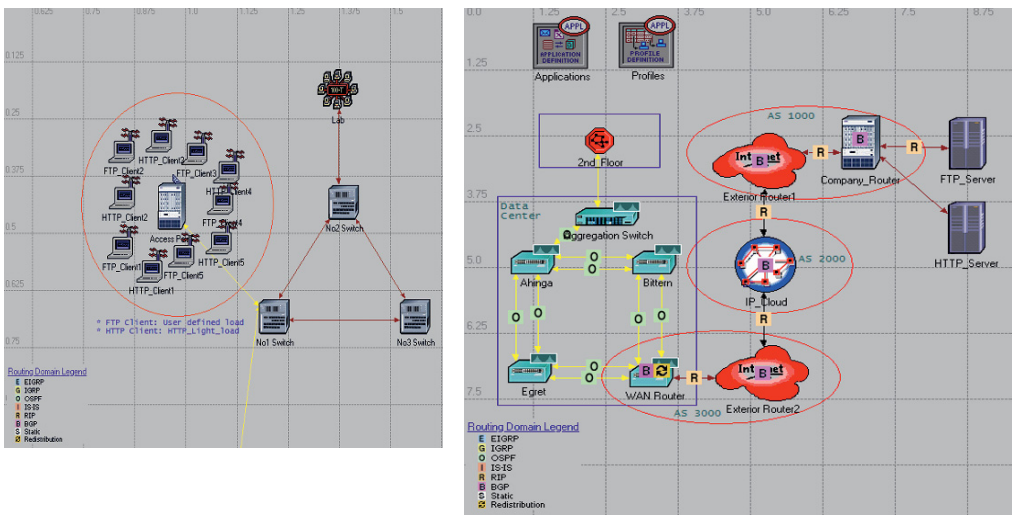


Figure 2. OPNET Simulation Model

B. OPNET Simulation Model

As illustrated in Figure 2, the model is comprised of three functional parts; *campus*, *Internet* and *service provider*. We model the campus data center with three LAN routers and a WAN router, and the building used in this deployment, UNF building 15, is modeled with one wireless access point and a maximum of ten wireless workstations connecting to the access point at any point of time.

The wireless workstations communicate with the access point, while the access point is connected to one of the Data Center LAN routers, through an aggregation switch. In fact, switches from all campus buildings are connected to the same aggregation switch in the data center. However in this model, we only consider the second floor of building 15 for the purpose of the simulation. The WAN router is connected to an exterior router.

The background traffic for this model is implemented as described below. The interior routing protocol used in the model is OSPF, while the choice for exterior routing protocol is BGP. Two exterior routers and an IP cloud are used to model the Internet. One exterior router is connected to a campus WAN router, while the other exterior router is connected to the proprietary router of the service provider mentioned above. They communicate with each other through the IP cloud, where the packet discard ratio is a controlled simulation parameter. The service provider was described having one router and two application servers FTP server and HTTP server. These two servers offer their services to the wireless workstations on the second floor of the campus building and to an Ethernet LAN of 10 workstations networked together.

C. Data Rates and other Simulation Parameters

The routers and the aggregation switch in the data center are connected by the 1000baseX lines, but it is the 100baseT between the access point and the router on second floor of the building. The supported data rate is 2 Mbps between the access point and wireless clients. The Interior Routing Protocol used on campus is OSPF.

The campus backbone network supports 1Gbps data rates but, the IP cloud (internet) between the exterior routers is connected by PPP DS3 link that supports only 44.736 Mbps. The BGP and RIP routing protocols are used among the routers from the WAN router in campus to the Internet service provider router. And each server of the Internet service provider is connected to the campus backbone network with 100 Mbps links. The RIP protocol is used between the ISP routers and each server.

IV. EVALUATION

We study the effect of differences in user Web browsing behavior on end user response times. We extend the work in [7], where OPNET is used as a simulation platform to model Web browsing behavior. This prior study models user accesses to Web documents of a median size of 7,328 bytes, with a total simulation duration of one hour. Our model is more realistic in that the mobile workstations and the WLAN's performance that we study are simulated with a background traffic generated by a multitude of other devices that depict an actual campus environment as described earlier in section III.

Our HTTP traffic characterization attempts are limited to temporal analysis [9], or the study of how frequently an event occurs. The collection of such statistics assumes that visitors can be uniquely identified and that requests can be traced in the sequence they occur. OPNET allows us to set individual browser requests with a mean of 300 seconds apart. Requests are pulled from an exponential distribution, similar to the work in [7]. OPNET does not provide the capability to track individual accesses to the actual Web documents accessed by clients and there is no support for application of *path analysis*, or the study of the sequence of page views by a user [9].

***Experiment 1*:**

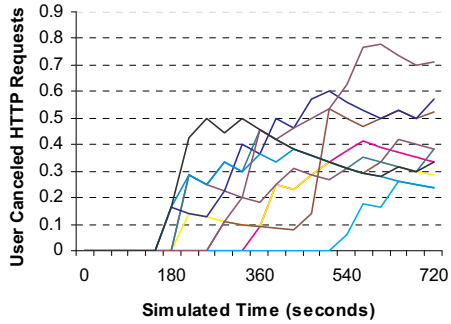
We measure end user canceled HTTP sessions and simulated end user response times, where the files sizes at an HTTP server are pulled from a heavy tailed Pareto distributed. The frequency of the request in Experiment 1 is 1 of 60 seconds per HTTP client. It has been shown before that Web page/object sizes are heavy tailed distributed [3]. We recognize that it has been shown before that transmission duration distributions for Web documents are heavy tailed [3] as well, and the distribution of file sizes has little effect on the end user transmission times distribution [3]. However, our results indicate, file size distribution has an effect on end user response time. This is why we choose the Web page document sizes to be pulled from Pareto distribution and not from uniform distribution, as used in the actual OPNET built-in wireless models.

The cumulative distribution function of the heavy tailed Pareto distribution is:

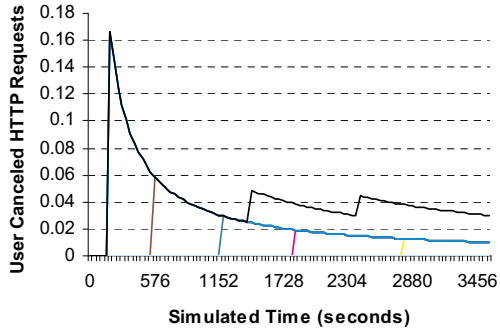
$$\Pr(X > x) = \left(\frac{x}{x_m}\right)^{-k},$$

where $x_m > 0$ is the minimum possible value of *X,* refereed to as *location parameter*, and $k > 0$ is a parameter referred to as the *Pareto index* or the *shape parameter*. For the purposes of this simulation, we choose a location parameter $x_m = 100{,}000$ (size of Web pages in bytes). File sizes follow Pareto distribution with shape parameter of $k = 1.1$, as found in [7] to realistically represent the distribution of Web page sizes.

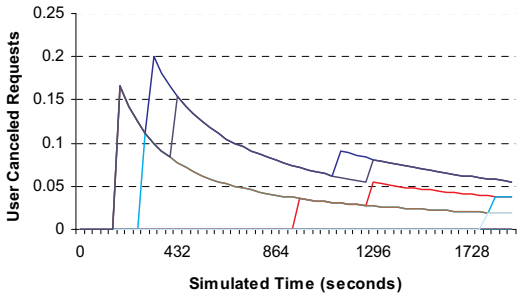
Mean interarrival times of Web page requests are modeled as exponential of 60 seconds. Total number of Web documents (pages) requested is 1000. Active TCP connections are about 3.8 on an average. This experiment resulted in an average of 0.31 Web page requests per second at the HTTP server. The results are shown in Figure 3 (c) and Figure 4 (c). Average number of user canceled connections is 0.014 and end user response time is 0.29 on an average for the duration of the simulation.



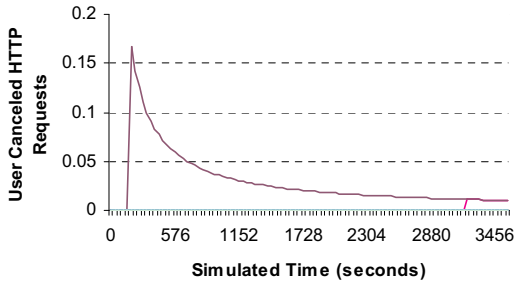
(a) Pareto Distributed Web page sizes of 1MBytes mean.



(c) Pareto Distributed Web page sizes of 10KBytes mean.



(b) Pareto Distributed Web page sizes of 100KBytes mean.



(d) Lognormall Web page sizes of 10KBytes mean.

Figure 3. Simulation Results: User Canceled HTTP Requests
(mean number of requests for each client are shown; colors are used to represent individual client results)

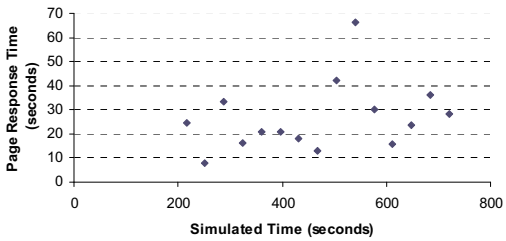
Experiment 2: We design this experiment to closely represent a similar experiment presented in [7]. We measure end user canceled HTTP sessions and simulated end user response times, where the files sizes at an HTTP server are modeled as lognormally distributed.

The cumulative distribution function of the lognormal distribution is:

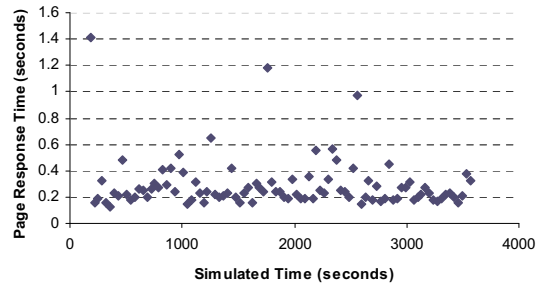
$$\Pr(X > x) = \frac{1}{2} + \frac{1}{2} erf[\frac{\ln(x) - \mu}{\sigma\sqrt{2}}],$$

where $x > 0$ and μ and σ are the mean and standard deviation of the normally distributed $\ln x$. We chose $\mu = 10{,}000$ and $\sigma = 9{,}290$, parameters that result in files size distribution similar to the one generated in [7], where file size minimum of 841 bytes, median of 7328 bytes, and maximum of 188,863 bytes has been used. This experiment resulted in an average of 0.31 Web page requests per second at the HTTP server. Active TCP connections are about 3.8 on an average. Results are shown in Figure 3 (d) and Figure 4 (d) and indicate that there is an average of 0.014 user canceled connections for the duration of the simulation, or the results of Experiment 1 constitute a 4.35 times increase in user canceled connections as compared to these results. End user delay is an average of 0.14 seconds per Web document. Or, the results of Experiment 1 constitute a 2.1 times increase in end user response time as compared to these results.

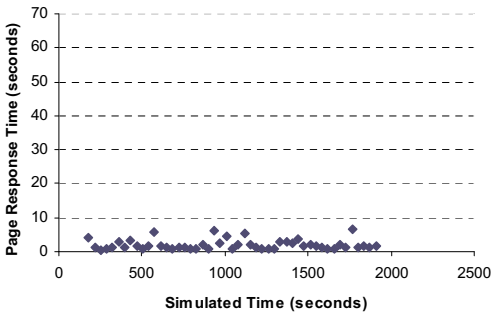
Experiment 3: We study the effect of heavy Web browsing vs. light Web browsing user profiles on end user response times. Our goal is to find a saturation point for a typical simple WLAN IEEE 802.11g network in a campus environment, where users start experiencing intolerable Web page response times, due to too much Web traffic generated within the WLAN.



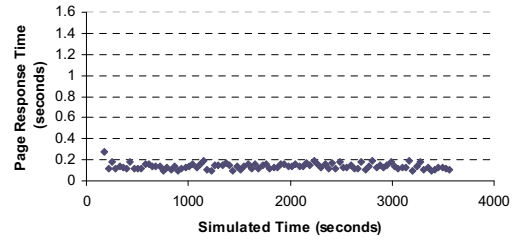
(a) 1Mbyte



(c) 10Kbyte



(b) 100 KByte



(d) Lognormal

Figure 4. Simulation Results: End User Response Time

We set the client maximum to 10 users, typical for a classroom environment and the frequency of the HTTP requests at an average of one every 60 seconds. Then, we vary the amount of traffic generated, by changing the sizes of requested Web pages. The rest of the parameters for this experiment are same as the parameters in *Experiment 1*. We closely monitor the simulated behavior of the HTTP server and all network components to ensure that no bottleneck is formed among them. For example, HTTP sever CPU utilization is below 1%. Average number of active TCP connections increases to greater than 10 with no server refused connections. The results are presented in Figure 3 (a) and (b) and Figure 4 (a) and (b).

V. Summary and Future Work

In this study we measured simulated end user response times. Our experiments were conducted to with files sizes at an HTTP server pulled from a heavy tailed Pareto distribution as well as lognormal distribution, as used in prior studies. Our results show that, *ceteris paribus*, response times when heavy tailed file sizes are modeled are two times worse than when modeling lognormal file sizes and studies that do not consider the results in [3] may present too optimistic results, not reflecting well real internet traffic.

The produced results also lead us to believe that even with a very small number of wireless users (10 clients in a typical campus small classroom environment) the end user delays become intolerable very fast with increasing the minimum HTTP document size.

Our overall conclusion, based on results from all experiments is that accesses to Web pages, containing large images and embedded documents, do not degrade end user response times in a 802.11g WLAN, as long as HTTP server directories contain a small number of such large files. That is, if file sizes are modeled by a heavy tailed Pareto distribution, where minimum file sizes are 100 Kbytes or less. When file size at the server become larger, with a minimum size of 1Mbytes, response times increase substantially and may become intolerable by humans, even if the number of WLAN users is not very high – a maximum of 10 users are connected in our model at any given point of time simulated for all experiments.

We outline as future work looking for trends in typical requested HTTP document sizes on the Web, similar to

Moore's Law for processor speed. We expect that these files sizes have increased in the last several years. We would like to not only quantify this increase, but more importantly look for trends since such trends can help researchers build more realistic simulation models over HTTP traffic.

VI. ACKNOWLEDGEMENTS

The authors would like to thank Steve Lyon and Daniel Simon with the University of North Florida's Information Technology services. They both helped the authors build a realistic simulation model based on the typical set up of networking equipment on the UNF campus.

REFERENCES

[1] A. Balachandran, G. Voelker, P. Bahl, and P. Rangan, in a Public Wireless LAN", Proceedings of ACM SIGMETRICS, Marina Del Rey, CA, pp. 195-205, June 2002.

[2] B. Bennington and C. Bartel, "Wireless Andrew: Experience Building a High Speed, Campus-WideWireless Data Network", Proceedings of ACMMOBICOM, Budapest, Hungary, pp. 55-65, September 1997

[3] M. Crovella, M. Taqqu, and A. Bestavros, "Heavy-Tailed Probability Distributions in the World Wide Web", book chapter from "A practical Guide To Heavy Tails: Statistical Techniques and Application", R. Adler, R. Feldman and M. S. Taqqu, Birkhauser, Boston, 1998.

[4] T. Hansen, P. Yalamanchili and H-W. Braun, "Wireless Measurement and Analysis on HPWREN", Proceedings of Passive and Active Measurement Workshop, Fort Collins, Co, pp. 222-229, March 2002

[5] IEEE 802.11 WIRELESS LOCAL AREA NETWORKS - The Working Group for WLAN StandardsURL: http://grouper.ieee.org/groups/802/11/

[6] D. Kotz and K. Essein, "Analysis of a Campus-Wide Wireless Network", Proceedings of ACM MOBICOM, Atlanta, GA, September 2002.

[7] Jan Libbenga, "Tokyo to enjoy city-wide WiMAX", Yozan MetroZone, March 2005. URL: http://www.theregister.co.uk/2005/03/18/tokyo-wimax/

[8] OPNET Technologies. URL: http://www.opnet.com

[9] James Pitkow, "In Search of Reliable Usage Data on the WWW", *Proceedings of the Sixth International Conference*, Palo Alto, CA, 1997.

[10] RFC 2616: "Hypertext Transfer Protocol – HTTP/1.1". URL: http://www.ietf.org/rfc/rfc2616.txt

[11] H. Singh and P. Singh, "Energy Consumption of TCP Reno, TCP NewReno, and SACK in Multihop Wireless Networks", Proceedings of A SIGMETRICS, Marina Del Rey, CA, pp. 206-216, June 2002

[12] D. Tang and M. Baker, "Analysis of a Local-Area Wireless Network", Proceedings of ACM MOBICOM, Boston, MA, pp. 1-10, August 2000.

[13] J. Yeo, M. Youssef, and A. Agrawala. A Framework for Wireless LAN Monitoring and its Applications. In Proceedings of the ACM Workshop on Wireless Security, pages 70ñ79, Philadelphia, PA, October 2004

Author Index

Subject Index