IJACSA

WHERE WISDOM SHARES

International Journal of Advanced Computer Science and Applications

Volume 2 Issue 8

August 2011

ISSN 2156-5570(Online)
ISSN 2158-107X(Print)

SAI

www.ijacsa.thesai.org

IJACSA

WHERE WISDOM SHARES

INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS

SAI

A Publication of
The Science and Information Organization

IJACSA Editorial

From the Desk of Managing Editor…

It is a pleasure to present our readers with the August 2011 Issue of International Journal of Advanced Computer Science and Applications (IJACSA).

The renaissance stimulated by the field of Computer Science is generating multiple formats and channels of communication and creativity. IJACSA is one of the most prominent publications in the field and engaging the ubiquitous spread of subject knowledge with effectiveness in all classes of audience. Nevertheless, the promise of increased engagement requires that we consider how this might be accomplished, delivering up-to-date and authoritative coverage of advanced computer science and applications.

The journal has a wide scope ranging from the many facets of methodological foundations to the details of technical issues and the aspects of industrial practice. It includes articles related to research findings, technical evaluations, and reviews. In addition it provides a forum for the exchange of information on all aspects.

The editorial board of the IJACSA consists of individuals who are committed to the search for high-quality research suitable for publication. These individuals, working with the editor to achieve IJACSA objectives, assess the quality, relevance, and readability of individual articles.

The contents include original research and innovative applications from all parts of the world. This interdisciplinary journal has brought together researchers from academia and industry as well as practitioners to share ideas, problems and solutions relating to computer science and application with its convergence strategies, and to disseminate the most innovative research. As a consequence only 31% of the received articles have been finally accepted for publication.

Therefore, IJACSA in general, could serve as a reliable resource for everybody loosely or tightly attached to this field of science.

The published papers are expected to present results of significant value to solve the various problems with application services and other problems which are within the scope of IJACSA. In addition, we expect they will trigger further related research and technological improvements relevant to our future lives.

We hope to continue exploring the always diverse and often astonishing fields in Advanced Computer Science and Applications.

Thank You for Sharing Wisdom!

Managing Editor
IJACSA
Volume 2 Issue 8, August 2011
editorijacsa@thesai.org
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

IJACSA Reviewer Board

CONTENTS

MCMC Particle Filter Using New Data Association Technique with Viterbi Filtered Gate Method for Multi-Target Tracking in Heavy Clutter

E. M.Saad
Department of Electronics and Communication Engineering
Helwan University
Cairo, Egypt

El. Bardawiny
Department of Radar
MTC College
Cairo, Egypt

H.I.ALI
Department of Electronics and Communication Engineering
Helwan University
Cairo, Egypt

N.M.Shawky
Department of Electronics and Communication Engineering
Helwan University
Cairo, Egypt

Abstract— **Improving data association technique in dense clutter environment for multi-target tracking used in Markov chain Monte Carlo based particle filter (MCMC-PF) are discussed in this paper. A new method named Viterbi filtered gate Markov chain Monte Carlo VFG-MCMC is introduced to avoid track swap and to overcome the issue of losing track to highly maneuvering targets in the presence of more background clutter and false signals. An adaptive search based on Viterbi algorithm is then used to detect the valid filtered data point in each target gate. The detected valid point for each target is applied to the estimation algorithm of MCMC-PF during calculating the sampling weights. This proposed method makes the MCMC interacts only with the valid target that is candidate from the filtered gate and no more calculations are considered for invalid targets. Simulation results demonstrate the effectiveness and better performance when compared to conventional algorithm MCMC-PF.**

Keywords- data association; multi-target tracking; particle filter; Viterbi algorithm ; Markov chain Monte Carlo; filtering

I. INTRODUCTION

In multi-target tracking systems (MTT) data association is an important key to associate the track to the true target in noisy measurements. The measurements (target of interest, clutter, noise signal) are reported by sensor at regular time interval (scan period). The received measurements in a cluttered environment may not originate from the real targets. Some of them may be from clutter or false alarm. As a result, the association between the tracked target and true candidate measurement is difficult. Assigning wrong measurements to track often results in lost track and track breaks. Moreover, in heavy clutter density the resulting number of false tracks may overwhelm the available computational resources of the MTT systems. For this reasons, improving data association technique has received much attention in MTT research [1-6]. The target tracking consists of two basic parts: data association and tracking filtering. Data association is responsible for detecting on each scan which of the received multiple measurements that lie in the specified gates of the predicted targets position should update with the existing tracking targets. Due to data association, each scan takes the output of the gating process [7] and makes final measurement to track associations. A gating process with filtering method based on Viterbi algorithm is used in this part to find the valid measurement that is considered as the best candidate measurement originated from the true target and assume the remaining measurements to be invalid that originated from false alarm or clutter. The Viterbi algorithm (VA) [8] is a dynamic programming technique for finding the shortest path through a trellis in a computationally efficient manner. The using of the Viterbi algorithm for solving the data association problem when tracking targets in clutter was first suggested as VA-QF by Quach and Farooq [9]. Their approach finds the approximate maximum likelihood assignment of measurements to the target. The problem they considered was that of tracking a single, maneuvering target assumed to be present in surveillance region. This method was modified in [10] to suit the problem of tracking multiple, non maneuvering targets in heavy clutter and was applied to tracking with an Over-The-Horizon radar (OTHR). Instead of using [9] in data association for single target that deals with multi state (multi-scan) where each state contains all measurements according to the corresponding time scan and is processed with the next scan up to all measurements received in the current scan. A new technique for gating process using Viterbi algorithm for multi-target tracking is introduced to be considered as an extension of [9] to detect the measurement in the current gate of the tracked target that has the maximum cost. This extension method for each track depends only on the measurements in current and previous gate of the same tracked target position by two successive states, each state represent the corresponding time scan for all tracked target gate. The candidate measurements in data association part are applied to tracking filter part for estimation. Due to the track swapping and highly maneuvering target, the mainly considered problem for estimation, traditionally, is solved using linearized filters, such as the extended Kalman filter (EKF) [11-15], under a

Gaussian noise assumption. The sufficient statistics from the linearized filter are used for data association. When dealing with non-linear models in state equation and measurement relation and a non- Gaussian noise assumption, these estimation methods may lead to non-optimal solutions. The sequential Monte Carlo methods, or particle filters [14-21], provide general solutions to many problems where linearizations and Gaussian approximations are intractable or would yield too low performance. Recently MCMC-based particle filter (MCMC-PF) [22-26] has captured the attention of many researchers in various applications that deal with difficult nonlinear and/or non Gaussian problems for tracking maneuvering target in clutter. In MCMC-PF, the particles are sampled from the target posterior distribution via direct MCMC sampling method, which avoids sample impoverishment. To increase the robustness of the algorithm and tracking a highly maneuvering target without swap in dense clutter environment, we introduce the proposed algorithm named VFG-MCMC. The new technique based on Viterbi algorithm as mentioned above is combined with MCMC-PF. Simulation results showed better performance and more effective at tracking when compared to the conventional MCMC-PF algorithm.

II. Problem Formulation

In a dynamic state space model, the observed signals (observation/ measurements) are associated with a state and measurement noise. Consider that there are *T* targets being tracked at time index *k*. The time evolution of the state $x_t(k) \in R^n$ of a discrete-time, nonlinear, dynamic systems is described by the following equation

$$x_t(k) = f(x_t(k-1), w_t(k-1)) \qquad t = 1,2,\dots,T \tag{1}$$

where *f* is the system transition function and $w_t(k-1) \in R^n$ is a dynamic noise which has a known probability density function (PDF). The superscript *t* corresponds to the t^{th} target. The initial target state, $x_t(0)$ for *t* = 1,2, ..., *T* , is assumed to be known. At discrete times, the measurements $z(k) \in R^m$ of the state $x_t(k)$ become available and are related to state through the observation equation

$$z(k) = h(x_t(k), v(k)) \qquad t = 1,2..T \tag{2}$$

where *h* is a nonlinear measurement function and $v(k) \in R^m$ is a sequence of observation noises of known PDF. From the Bayesian perspective, the tracking problem is to recursively calculate the posterior probability density function, $p(x_t(k) \mid z_{1:k})$, where $z_{1:k} = \{z_m,\ m = 1,\dots..k\}$.

A. Viterbi Data Association VA-QF for Single Target

The basis of the VA algorithm is built upon the Markov process [9],[27],[28]. In VA algorithm, the target motion is assumed to be a Markov process. The Markov process uses state diagrams where the nodes or measurements represent the state in each scan and the arcs are represented by transition paths. As time progress, the process will trace some path from state (or scan) to state through the state diagram. If the states and their transitions are repeated in time for k=1,..,K, we obtain a trellis that represent a directed graph consisting of a set of measurements and paths as shown in Fig. 1. At each measurement from scan to scan, the length of the paths coming into the measurement is summed. The path with the shortest length will be assigned to that measurement and is considered as survivor path (one survivor for each measurement). the recursion will proceed from first to last (current) scan, once the last measurement is reached, the algorithm summing all the smallest (path metric) nodes from each previous scan to obtain the optimal path. A path is a collection of directed arcs that connects an element at stage 1 to an element at stage K. Each directed arcs is associated with a metric or a distance label $a_{ij}(k)$. A path metric is defined as the sum of the metrics of all the arcs contained in the path as

$$d(p) = \sum_{k=2}^{K} a_{ij}(k); \qquad \{n_i(k-1), n_j(k)\} \in p$$

(3)

By applying the Viterbi algorithm to the trellis diagram, we obtain the sequence of measurements that maximizes the overall likelihood function, thus solving the data association problem as in the following steps.

1) Initialization step: assign a value of zero to the label at each node/measurement in stage 1:

$$score_i(1) = 0, \qquad 1 \le i \le n_1$$

$$\psi_i(1) = 0, \qquad 1 \le i \le n_1$$

2) Repeat the following procedure for each stage *k*, where *k=2,…K*

- For each node *i*=0,….. n_{k-1} (at stage *k-1*) , calculate the predicted position $\hat{z}_i(k \mid k-1)$ conditioned on the measurement sequence termination at the node $n_i(k-1)$
- For each node *j*=0,….. n_k (at stage *k*) , calculate the distance label $a_{ij}(k)$ by calculating the normalized negative log likelihood function (non –maneuvering target) of the arc joining nodes $n_i(k-1)$ and $n_j(k)$ as given in [27-29]

$$a_{ij}^{k} = \left\{\begin{array}{ll} -\ln\{\Lambda(.)\}, & j \ne 0, i = 1,\dots., n_{k-1} \\ 0 & j = 0, i = 1,\dots. n_{k-1} \end{array}\right\}$$

$$-\ln\{\Lambda(.)\} = a_{ij}(k) = \left[0.5(\tilde{z}_{ij}(k))'(s_i(k))^{-1}(\tilde{z}_{ij}(k))\right] \\ -\ln\left\{\frac{p_D V_s}{1-p_D}\right\} + \ln\left\{(2\pi)^{n/2}\left|s_i(k)\right|^{0.5}\right\} \tag{4}$$

where the innovation $\tilde{z}_{ij}(k) = z_j(k) - \hat{z}_i(k \mid k-1)$

- assign node $n_j(k)$ with the smallest label as follow:

$$d_j(k) = \min_{0 \le i \le n(k-1)} \{d_i(k-1) + a_{ij}(k)\} \qquad (5)$$

$$\overset{*}{i} = \arg\left\{ \min_{0 \le i \le n(k-1)} \{d_i(k-1) + a_{ij}(k)\} \right\} \qquad (6)$$

and let

$$score_j(k) = (d_i(k-1) + a_{ij}(k)) \qquad (7)$$

$$\psi_j(k) = \overset{*}{i} \qquad (8)$$

- update the target state $\hat{x}_j(k)$ and its covariance $p_j(k)$ at each node in stage k

3) Final selection: determine the node with the minimum score in the final stage. The current state of the target is the state associated with the minimal node.

$$\overset{*}{i} = \arg\left\{ \min_{0 \le i \le n_K} \{score_i(k)\} \right\} \qquad (9)$$

$$\hat{x}(K) = \hat{x}_{\overset{*}{i}}(K) \qquad (10)$$

4) Back track: recover the measurement sequence that terminates with the minimum node score in the final stage using

$$\overset{*}{i}|_{k-1} = \psi_{\overset{*}{i}}(k) \qquad (11)$$

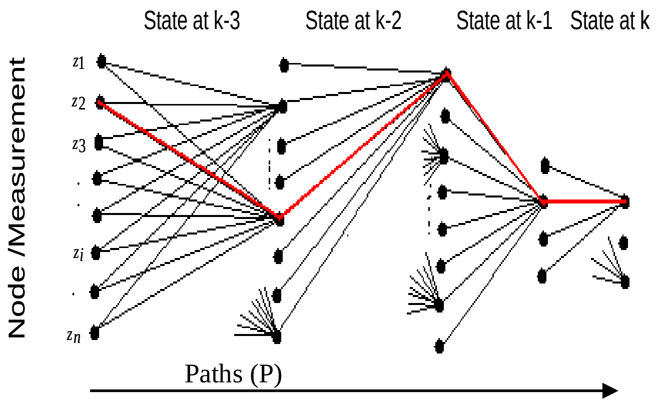


Figure 1. Trellis Diagram of VDA for tracking target

B. Markov Chain Monte Carlo based Particle Filter(MCMC -PF)

1) Basic theory of particle filter

The particle filter is a recent filtering method based on Bayesian estimation and Monte Carlo method and effectively deals with complicated nonlinear and /or non-Gaussian problems. The basic idea of this method is Monte Carlo simulation, in which the posterior density is approximated by a set of particles with associated weights $\left\{x_t^i(k-1), w_t^i(k-1) \quad i = 1,2,..N\right\}$.

At every time step we sample from the proposal distribution $q(x_t(k)/x_t(k-1), z_{1:k})$ to achieve new particles and compute new weights according to the particle likelihoods. After normalization of weights, the posterior density can be represented by $\left\{x_t^i(k), w_t^i(k) \quad i = 1,2,..N\right\}$.

The implementation of the basic particle filter is as follows:

(1) Initialization

Draw a set of particles from the prior $p(x_t(0))$:

$$x_t^i(0) \sim p(x_t(0)), \quad i = 1,2\ldots,N \qquad (12)$$

Prediction:

(2) Sampling step

(a) for $i=1,2,\ldots N$

Sample $x_t^i(k)$ from the proposal distribution $q(x_t(k) \mid x_t(k-1), z_{1:k})$:

$$x_t^i(k) \sim q(x_t(k) \mid x_t(k-1), z_{1:k}) \qquad (13)$$

Update:

(b) Evaluate the importance weights when the candidate measurement $z(k)$ is choosing by data association algorithm

$$w_t^i(k) = w_t^i(k-1) \frac{p(z(k) \mid x_t^i(k))\, p(x_t^i(k) \mid x_t^i(k-1))}{q(x_t(k) \mid x_t(k-1), z_{1:k})} \qquad (14)$$

(c) Normalize the weights

$$\tilde{w}_t^i(k) = w_t^i(k) / \sum_{i=1}^{N} w_t^i(k) \quad ,i = 1,2\ldots,N \qquad (15)$$

(3) Markov Chain Monte Carlo (MCMC) step as described in algorithm 1

(4) Output step

Output a set of particles $\left\{x_t^i(k), w_t^i(k) \quad i = 1,2,..N\right\}$ that can be used to approximate the posterior distribution.

Expectation:

$$\hat{x}_t(k) = \sum_{i=1}^{N} \tilde{w}_t^i(k)\, x_t^i(k) \qquad (16)$$

Covariance:

$$p_t(k) = \sum_{i=1}^{N} \tilde{w}_t^i(k)(x_t^i(k) - \hat{x}_t(k))(x_t^i(k) - \hat{x}_t(k))' \qquad (17)$$

and set $w_t^i(k)=1/N, \quad i=1,2,..N$ (18)

(5) k=k+1, go to step 2

2) Basic theory of Markov Chain Monte Carlo (MCMC)

The MCMC step, as described in [24] has an invariant distribution $\prod_{i=1}^{N} P(x_{t\;0:k}^{i} \mid z1:k)$, which is applied to each of the N particles, one at the time.

The metropolis- Hastings algorithm (MH) is a way to simulate from such a chain. The idea of MH is to sample states from a Markov Chain with the posterior as invariant distribution.

A Markov Chain is constructed by approximate a candidate for the next state $x_t^{*i}(k)$ given the current state $x_t^i(k)$ according to the proposal distribution $p(x_t(k) \mid x_t(k-1))$. This state transition is accepted with probability

$$\alpha(x_t^i(k), x_t^{*i}(k)) = \min\left\{1, \frac{p(z(k) \mid x_t^{*i}(k))}{p(z(k) \mid x_t^i(k))}\right\} \qquad (19)$$

Algorithm 1: Markov Chain Monte Carlo (MCMC) Step

(a) For each particle 1:N sample the proposal candidate for the approximation of the next state.

$$x_t^{*i}(k) \sim p(x_t(k) / x_t(k-1))$$

(b) Sample $\rho \sim$ U(0,1),where U(0,1) is a uniform distribution in the interval (0,1).

(c) If $\rho \le \min\left\{1, \frac{p(z(k) \mid x_t^{*i}(k))}{p(z(k) \mid x_t^i(k))}\right\}$

then accept move:

$$x_t^i(k) = x_t^{*i}(k)$$

else reject move

$$x_t^i(k) = x_t^i(k)$$

end if.

III. VITERBI FILTERED GATE DATA ASSOCIATION

To introduce a robust filtering technique with improving tracking performance and give the tracking system the capability to track the multi-target when moving with highly maneuvering without swapping and in dense clutter environment, a new algorithm, named Viterbi filtered gate based Markov Chain Monte Carlo- particle filter (VFG-MCMC) as shown in Fig. 2, is proposed.

In the prediction step, Let $Z(k-1)=\{z_1(k-1), z_2(k-1), \dots z_{w_n}(k-1)\}$ be a set of points in the 2-D Euclidean space at time k-1 where w_n is the number of points at time scan Δt and let $\hat{z}^t(k)$ be a predicted position of the t^{th} tracked target at time k. according to distance metric measure and gate size, let $\bar{Z}^t(k-1)=\{z_1(k-1), ..z_i(k-1), ..z_{m_i}(k-1)\}$ be a set of the candidate points detected in the t^{th} gate $G_t(k-1)$ of predicted position $\hat{z}^t(k)$ whose elements are a subset from the set $Z(k-1)$ where i =1 to mi (number of detected points in gate $G_t(k-1)$ at time k-1) and $\bar{Z}^t(k-1)$ be a set of all valid points $z_i(k-1)$ that satisfy the distance measure condition

$$\left(z_i(k-1)-\hat{z}^t(k)\right)' S^t(k)^{-1}\left(z_i(k-1)-\hat{z}^t(k)\right) \le \gamma \qquad (20)$$

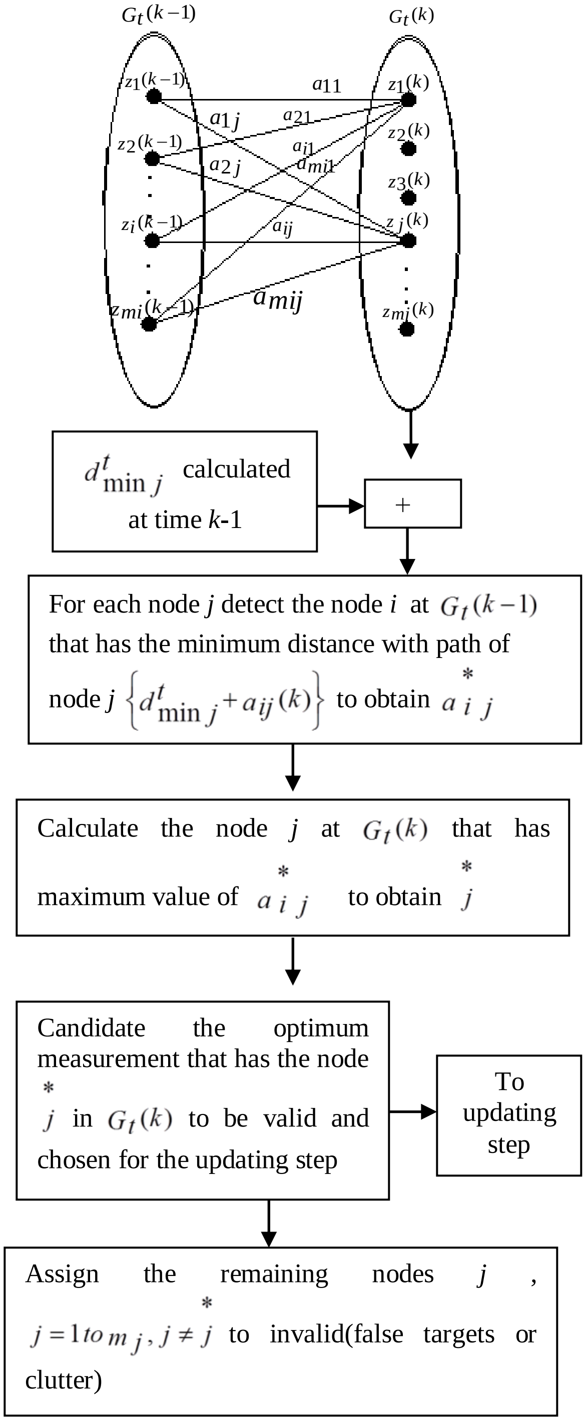
for each target t where γ is threshold value that determines the gate size and l =1 to w_n , i =1 to mi, i. e for each target t, i is initialized by 1 and is increased by $i = i$ +1 after each valid point is detected up to last mi detected points.

In the updating step, let $Z(k)=\{z_1(k), z_2(k), \dots z_{w_c}(k)\}$ be a set of points in the 2-D Euclidean space at time k where w_c is the number of points at time scan Δt. The candidate points detected in the same gate $G_t(k)$ as $G_t(k-1)$ of the t^{th} predicted position $\hat{z}^t(k)$ is a subset $\bar{Z}^t(k)=\{z_1(k), ..z_j(k), ..z_{mj}(k)\}$ from the set $Z(k)$ where j =1 to mj (number of detected points in t^{th} gate at time k) and $\bar{Z}^t(k)$ be a set of all valid points $z_j(k)$ that satisfy the distance measure condition

$$\left(z_j(k)-\hat{z}^t(k)\right)' S^t(k)^{-1}\left(z_j(k)-\hat{z}^t(k)\right) \le \gamma \qquad (21)$$

for each target t where l =1 to w_c , j=1 to mj for j=j+1 after each valid point is detected. A new filtering gate method based Viterbi algorithm is proposed to distinguish between the detected measurements in $G_t(k)$ that originated from the target or originated from clutter (false target).

In this method we consider the two constructed gate $G_t(k-1), G_t(k)$ for each target as the two successive state as mentioned above in viterbi algorithm VA-QF for single target. This method is based only on the measurements (nodes) $\{z_i(k-1), z_j(k)\}$that fall in the two previous and current gate $\{G_t(k-1), G_t(k)\}$ for each target that have the same predicted target position at time scan k-1, k respectively.



For each node j detect the node i at $G_t(k-1)$ that has the minimum distance with path of node j $\left\{d^t_{\min j}+a_{ij}(k)\right\}$ to obtain $a_{\overset{*}{i}j}$

Calculate the node j at $G_t(k)$ that has maximum value of $a_{\overset{*}{i}j}$ to obtain $\overset{*}{j}$

Candidate the optimum measurement that has the node $\overset{*}{j}$ in $G_t(k)$ to be valid and chosen for the updating step

Assign the remaining nodes j, $j=1\,to\,m_j, j\neq\overset{*}{j}$ to invalid(false targets or clutter)

Figure 2. Schematic Diagram of VFG data association algorithm

To detect the candidate valid measurement $z_{\overset{*}{j}}(k)$ from the set $\bar{Z}^t(k)$ of the t^{th} gate $G_t(k)$ to be considered as the valid measurement that originate from the target and as the choosing measurement in the updating step, we first calculate distance metric between each measurement $z_j(k)$ in current gate $G_t(k)$ and all measurements $z_i(k-1)$ in its previous gate $G_t(k-1)$ as given

$$a^k_{ij}=\begin{cases}-\ln\{\Lambda(.)\}, & j=1,\ldots m_j, i=1,\ldots,m_i\\ 0 & j=0, i=1,\ldots m_i\end{cases}\qquad(22)$$

$$-\ln\{\Lambda(.)\}=a_{ij}(k)=\left[0.5(\tilde{z}_{ij}(k))'(s_i(k))^{-1}(\tilde{z}_{ij}(k))\right]-\ln\left\{\frac{p_D V_s}{1-p_D}\right\}+\ln\left\{(2\pi)^{n/2}|s_i(k)|^{0.5}\right\}\qquad(23)$$

where the innovation in this method is calculated as

$$\tilde{z}_{ij}(k)=z_j(k)-z_i(k-1)\qquad(24)$$

The minimum distance $d^t_{\min j}$ that represents the shortest path at each measurement $z_j(k)$ is assigned to its corresponding distance metric $a_{ij}(k)$ to give

$$a^+{}_{ij}(k)=d^t_{\min j}+a_{ij}(k),\quad j=1,\ldots m_j, i=1,\ldots,m_i\qquad(25)$$

For each measurement label j in $G_t(k)$, we choose one of the path metric $a^+{}_{ij}(k)$ $,i=1,\ldots,m_i$ that has the minimum distance and then its measurement label (index) $\overset{*}{i}$ in $G_t(k-1)$ is obtained by

$$\left.\begin{aligned}&\overset{*}{i}=\arg\left\{\min_{1\le i\le m_i}\left\{a^+{}_{ij}(k)\right\}\right\},\\ &score_j(k)=a_{\overset{*}{i}j}\\ &d^t_{\min j}=d^t_{\min j}+a_{\overset{*}{i}j}\end{aligned}\right\}\ j=1,\ldots m_j\qquad(26)$$

After determining the shortest path with minimum distance $a_{\overset{*}{i}j}$ for each measurement label j in $G_t(k)$, the valid measurement $\overset{*}{j}$ that is a candidate to be the associating measurement to a tracked target t in the updating step is calculated by obtaining the measurement index label of the maximum $score_j(k)$ as in the following equation

$$\overset{*}{j}=\arg\max_{1\le j\le m_j} score_j(k)\qquad(27)$$

Thus, for each target *t* at each new scan, the filtered gate using this proposed method of data association technique select one measurement from the target gate that is considered to be valid (originate from the true target) and is assigned to be the selected or candidate measurement used in the process for estimation in the updating step, while the remaining measurements are considered to be invalid (originate from clutter) and are avoided from tracking process.

IV. Implimentation of VITERBI Filtered Gate Data Association with MCMC–PARTICLE FILTER (VFG-MCMC)

The proposed algorithm VFG-MCMC is an approximate likelihood- based approach for solving the data association problem in multi target tracking when the clutter is heavy. This new algorithm depends on the history of measurements for one scan and the construction current gate. The current gate of the tracked target is processed with the previous gate of the same tracked target position using an advanced technique based on the basic principle of viterbi algorithm VA-QF. This technique detects the one valid measurement in the current gate that is assumed to be originated from true target, while the remaining measurements in the gate are considered as invalid that are assumed to be originated from clutter or false target. This filtering gate method select or candidate the one valid measurement for each target to be processed in the updating step while the invalid measurements not include in the tracking process. The proposed data association algorithm is combined with MCMC-PF to overcome the problem of highly maneuvering target and swapping in the presence of dense clutter. The proposed algorithm VFG using MCMC-PF is represented in algorithm 2.

Algorithm 2: VFG-MCMC

For t=1 to T

Initialization

1- Draw a set of particles from the prior $p(x_t(0))$:

$x_t^i(0) \sim p(x_t(0)), \quad i=1,2\ldots,N \qquad , \qquad w_t^i(0)=1/N$

Prediction:

2- With sampling step *Particles at time step k*-1, $\{\hat{x}_t^i(k-1)\}$, are passed through the dynamic state model as in (1) to obtain the predicted particles

at time step *k*, $\{\hat{x}_t^i(k)\}$:

$$\hat{x}_t^i(k) = f(\hat{x}_t^i(k-1), \omega_t^i(k-1)) \qquad t=1,2,\ldots,T \text{ , } i=1,2,\ldots N$$

3- Without sampling step

(a)Expectation *at time step k*-1, $\{\hat{x}_t(k-1)\}$, are passed through the dynamic state model as in (1) to obtain the predicted particles at time step *k*, $\{\hat{x}_t(k)\}$:

$$\hat{x}_t(k) = f(\hat{x}_t(k-1), \omega_t(k-1)) \qquad t=1,2,\ldots,T$$

$$\hat{z}^t(k) = h(\hat{x}_t(k), \qquad t=1,2..T$$

(b) Finding validate measurement $z(k) = \overset{*}{z}_j(k)$ according to algorithm 3

Update:

4- Once the measurement data $z(k)$ is obtained by data association algorithm VFG, evaluate the importance weight of each predicted particle

$$w_t^i(k) = w_t^i(k-1)\, p(z(k) \mid \hat{x}_i^i(k))$$

and normalize the weights

$$\tilde{w}_t^i(k) = w_t^i(k) / \sum_{i=1}^{N} w_t^i(k) \quad , i=1,2\ldots,N$$

5- Markov Chain Monte Carlo (MCMC) step as described in algorithm 1

6- Output step

Output a set of particles $\{x_t^i(k), w_t^i(k) \quad i=1,2,..N\}$ that can be used to approximate the posterior distribution by Calculating the expectation $\hat{x}_t(k)$ and Covariance $p_t(k)$ using (16),(17) respectively

and set $w_t^i(k)$ as in (18)

7- *k*=*k*+1, go to step 2

8- End for

Algorithm 3: Find the validate measurement using VFG algorithm

1. Find validated region for measurements at time *k*-1:

$$\overline{Z}^t(k-1) = \{z_i(k-1)\}, \qquad i=1,..mi$$

By accepting only those measurements that lie inside the gate *t*:

$$\overline{Z}^t(k-1) = \left\{ Z \; : \left(z_i(k-1) - \hat{z}^t(k)\right)' S^t(k)^{-1} \left(z_i(k-1) - \hat{z}^t(k)\right) \le \gamma \right\}$$

2. Find validated region for measurements at time *k*:

$$\bar{Z}^{t}(k) = \{z_j(k)\}, \qquad j = 1,..mj$$

By accepting only those measurements that lie inside the gate t

$$\bar{Z}^{t}(k) = \left\{ Z \; : \left(z_j(k) - \hat{z}^t(k)\right)' S^t(k)^{-1} \left(z_j(k) - \hat{z}^t(k)\right) \le \gamma \right\}$$

3. For each measurement *j* the distance metric a_{ij}^{k} is calculated by (22) using (23), (24)

4. The distance path $a^{+}_{ij}(k)$ is obtained by (25)

5. For each measurement *j* finds the label measurement $\overset{*}{i}$ as calculated in (26) that has the minimum distance $a_{\overset{*}{i}j}$

6. Find the valid measurement $\overset{*}{j}$ that has the maximum score Using (27).

7. The associated measurement $z(k)$ is set to

$$z(k) = z_{\overset{*}{j}}(k)$$

Finally, by using VFG data association algorithm ,we obtain for each target the valid measurement in the tracked target gate that is applied to MCMC-PF before entering in the process of the updating step . The other measurements in the gate are assumed to be invalid and the updating process for tracking the targets is not assigned to any one of them. VFG data association algorithm has an advantage that is accepted by MCMC-PF to improve the performance in heavy clutter for the tracked targets and to maintain the tracking to the targets that move with high maneuvering and swapping.

V. Simulation Results

In order to demonstrate the performance of the proposed VFG-MCMC algorithm, we provide a comparison with the most popular nonlinear conventional tracking algorithm MCMC-PF. Two test scenarios have been chosen in this section for moving more than one target in the XY plane with different issues according to the two different tracking problems; high maneuvering and swapping in dense clutter environment. The discrete time state equation with sampling interval Δt = 4 sec is

$$x_t(k) = F\, x_t(k-1) + \Gamma\, w_t(k-1) \qquad t = 1,2,.T$$

where

$x_t(k) = [x(k), y(k), \dot{x}(k), \dot{y}(k)]$ is the state vector ; $x(k)$ and $\dot{x}(k)$ are respectively the position and velocity of the moving object along the Cartesian frame X axis; and $y(k)$, $\dot{y}(k)$ along the Y axis. Γ is a unity matrix and $w_t(k-1)$ is discrete time white, Gaussian noise: $w(k) \sim N(0, Q_k)$, $Q = G \;\; G'$. The measurements are received from one sensor which is positioned at the origin of the plane. The measurement equation is as follow:

$$z(k) = h(x_t(k)) + v(k) \qquad t = 1,2..T$$

, where $z(k) = [z_1(k), z_2(k)]$ is the observation vector. $z_1(k)$ is the distance between the origin and the moving target and $z_2(k)$ is the bearing angle. The measurement noise $v(k) = [v_{z1}(k), v_{z2}(k)]$ is a zero mean Gaussian white noise process with variance R , where $R = \begin{bmatrix} \sigma_{z1}^2 & 0 \\ 0 & \sigma_{z2}^2 \end{bmatrix}$,

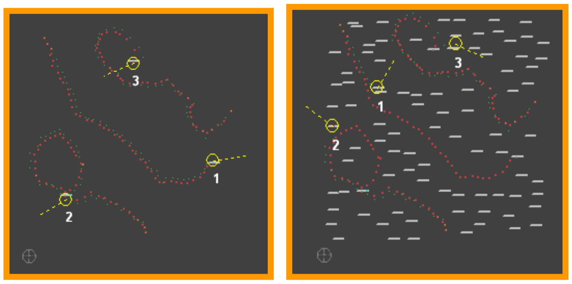
$$z(k) = \begin{bmatrix} \sqrt{(x(k) - x_r)^2 + (y(k) - y_r)^2} + v_{z1}(k) \\ \arctan\left(\dfrac{y(k) - y_r}{x(k) - x_r}\right) + v_{z2}(k) \end{bmatrix}$$

where (x_r, y_r) is the position of the sensor at the origin,

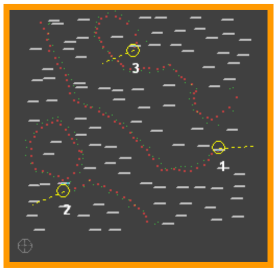
$$F = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad G = \begin{bmatrix} \dfrac{\Delta t^2}{2} & 0 \\ 0 & \dfrac{\Delta t^2}{2} \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix}$$

The first test scenario has three tracks with initial state $x_1(0)$ = [13.4km, 7.15km, 200m/s, 200m/s], $x_2(0)$ = [16.1km, 12.3km, 200m/s, 200m/s], $x_3(0)$ = [18.3km, 9.09km, 200m/s, 200m/s], which continues from the first frame to the last frame. This scenario is used to evaluate the tracking performance by the proposed and conventional algorithm when the targets move with high maneuvering in the presence of heavy clutter. The second test scenario has two tracks with initial state $x_1(0)$ = [13.4km, 12.5km, 150m/s, 150m/s], $x_2(0)$ = [13.3km, 10.5km, 150m/s, 150m/s]moving in cross trajectory corrupted by more background of clutter to evaluate the tracking performance against swapping between tracks for the two approaches that are presented here. We initiate the other parameters as; the row and column sizes of the volume ($V_S = S_W \times S_H$), V_S =20x20, $T = 38*4 = 152$ sec, p_D =0.99, in addition to $R = \begin{bmatrix} 400m^2 & 0 \\ 0 & 1\deg^2 \end{bmatrix}$, the particle number *N*=200,sensor position (x_r, y_r)= (0,0)

According to the first test scenario, using the conventional MCMC-PF algorithm in clear environment with no clutter to track highly maneuvering targets, continuing to tracking with no failure as expected as shown in Fig.3 (a). The X-Y trajectory is implemented in Fig 4(a). But this algorithm at low signal to noise ratio fails to track the corrupted targets with a uniform high clutter density as shown in Fig. 3(b) while, given a fixed threshold (γ $=10^{-4}$), we showed that the proposed VFG-MCMC succeeded to track the targets in the same environment as shown in Fig. 3(c). We obtain trajectories for X- and Y- components as shown in Fig. 4(b),(c). In these figures, the colored solid line represents the underlying truth targets of the trajectory (each target with different color) while the colored + symbol represents trajectory of the estimated tracked targets. Our proposed algorithm (+ symbol with different color) detects and associates the proper sequence of observation very well compared to MCMC-PF which fails to continue. According to the second test scenario, the conventional algorithm in clear environment with no clutter fails to avoid swapping between tracks as shown in Fig. 5(a),(b),(c),(d) that are represented in different time scan and also its X- Y- trajectory is implemented as shown in Fig. 6 . Thus, when using the conventional algorithm for tracking the targets in more background clutter we fail to track the targets with no avoiding the swapping of the crossed targets as shown in Fig. 7(a). The proposed algorithm that is evaluated in the same clutter environment succeeded to track the targets with no swapping as shown in Fig.7 (b). The X- Y- trajectories for the two approaches to demonstrate the swapping effect and the capability to maintain the tracks in more background clutter are shown in Fig. 8 (a),(b). According to the first and second test scenario we also compared error root mean square value (RMSE) for the different two approaches with three targets and with two targets as shown in Fig. 9,10. Our proposed algorithm in dense clutter environment has lower error, RMSE values and less sensitive to clutter than MCMC-PF over frame numbers.

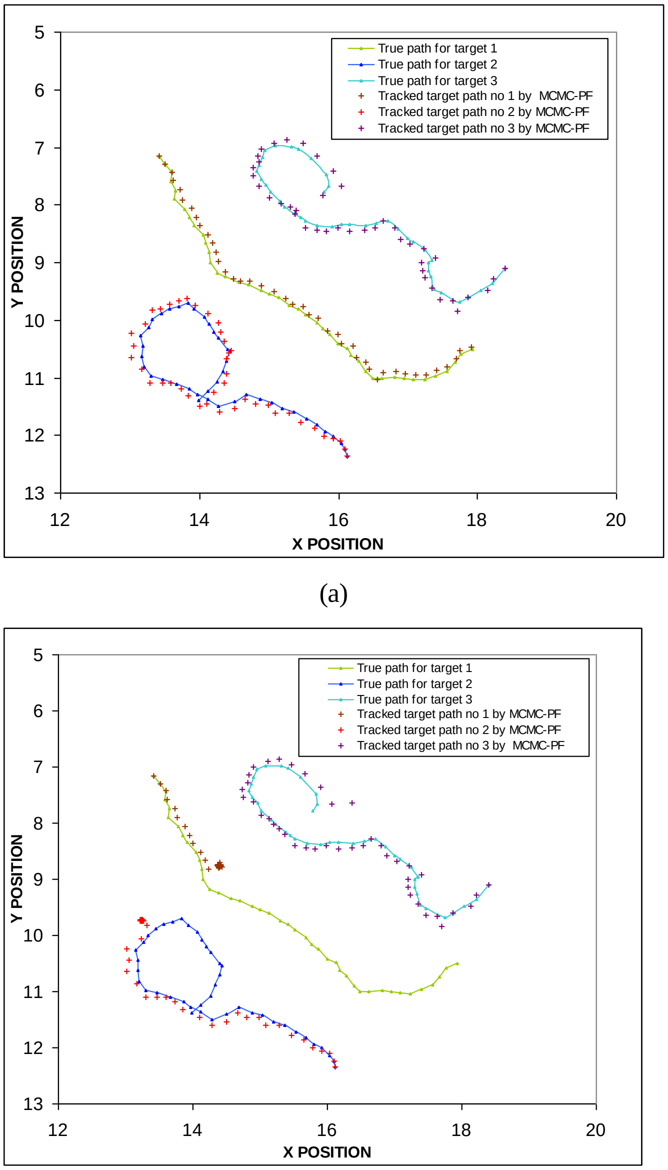


(a) (b)



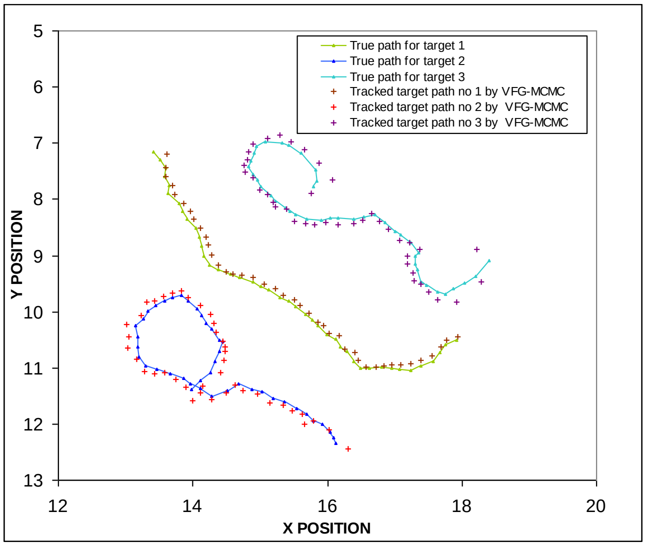
(c)

Figure 3. The state of tracking multi-targets (3 moving targets) using the 2 approaches (a) MCMC-PF succeeded to track in clear environment (no clutter) (b) MCMC-PF failed to continue tracking in heavy clutter (c) VFG-MCMC succeeded to continue tracking in the same environment as in (b)



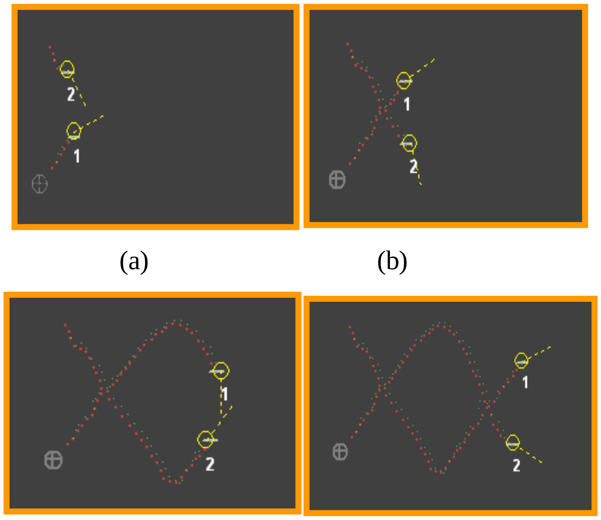
(a)

(b)



(c)

Figure 4. Trajectory for X-Y components for the MCMC-PF used in tracking 3 targets (+ symbol) without background clutter as in (a) and for the 2 approaches algorithm with heavy clutter as in (b),(c). all showed the true targets path as in solid line. (b) Trajectory for X-Y by MCMC-PF. (c) Trajectory for X-Y by VFG-MCMC.



(a) (b)

(c) (d)

Figure 5. The swapping state for tracking 2 moving targets in cross trajectory using MCMC-PF at different time scan with no clutter as in (a),(b),(c),(d) This show the track no. 1 swap with track no..2 as shown in (d)

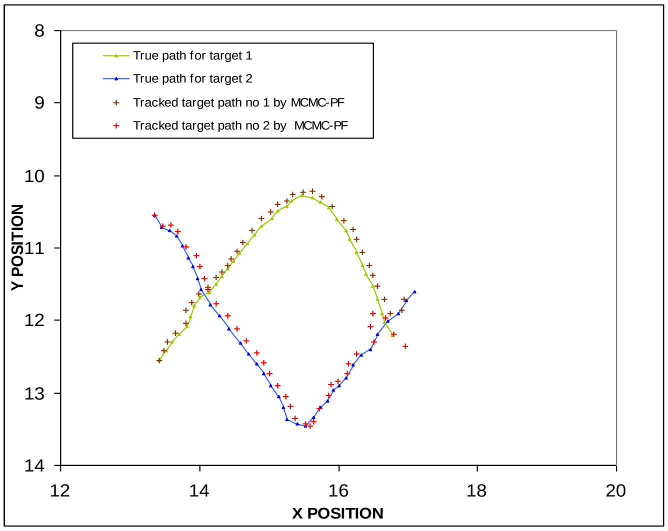
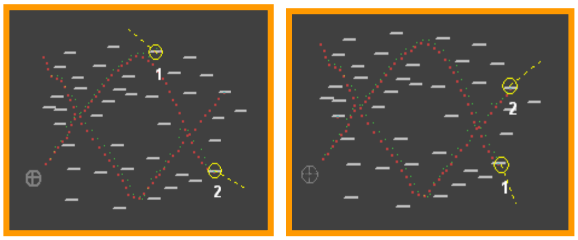
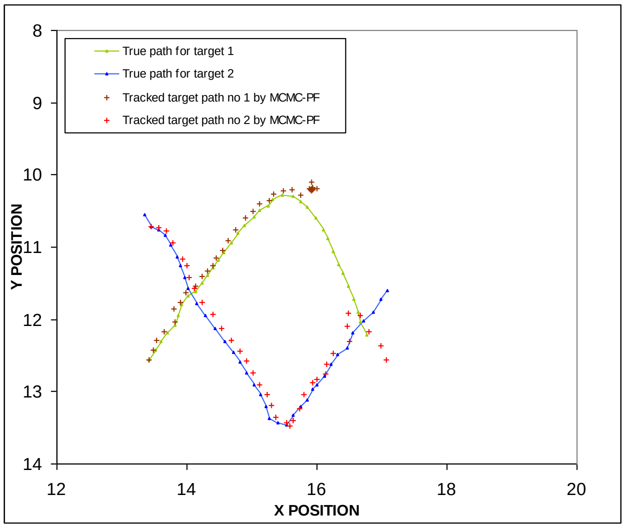


Figure 6. X-Y Components Trajectory for the MCMC-PF used in tracking 2 targets (+ symbol) without background clutter and the true targets Path in solid line. This showed the MCMC-PF fails to avoid swapping.

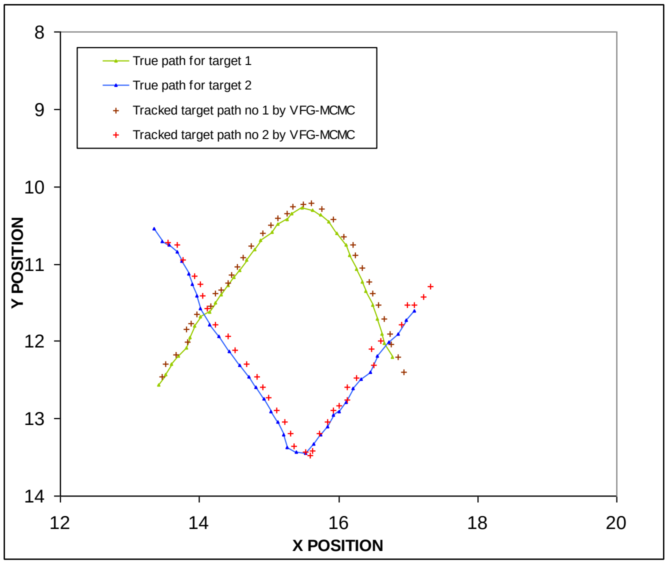


(a) (b)

Figure 7. The swapping state for tracking 2 moving targets in cross trajectory using the 2 approaches in more background clutter (a) MCMC-PF fails to continue to track no 1 and take the wrong trajectory for track no 2 (b) VFG- MCMC succeeded to continue tracking and avoided swapping .



(a)



(b)

Figure 8. X-Y Components Trajectory for the 2 approaches algorithm used in tracking 2 targets (+ symbol) in more background clutter and showed the true target path as in solid line. (a) Trajectory for X-Y by MCMC-PF (b) Trajectory for X-Y by VFG-MCMC.

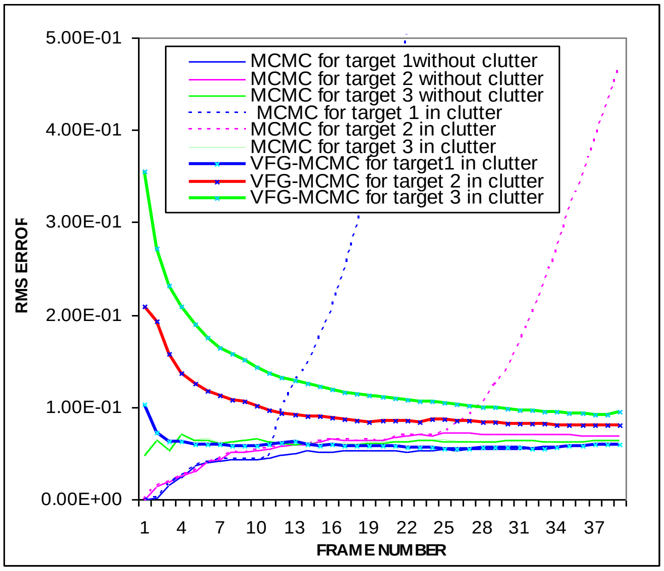


Figure 9. The RMSE for each target (3 targets) separately over frame number (each frame take 4 sec / one scan) according to tracking evaluation of the first test scenarios as shown in Fig. 3,4 .

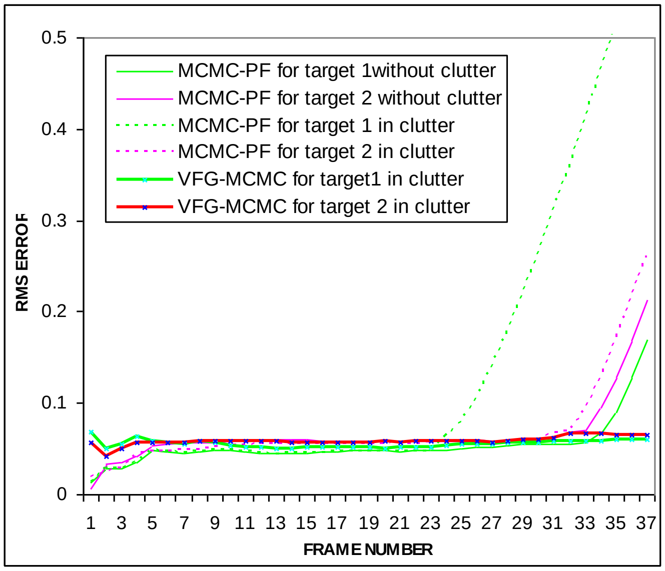


Figure 10. The RMSE for each target (2 targets) separately over frame number (each frame take 4 sec / one scan) according to tracking evaluation of the second test scenarios as shown in Fig. 5,6,7,8.

VI. Conclusion

In this paper for multi-target tracking in dense clutter environment, a new data association technique based on filtering gate method using Viterbi algorithm (VFG) applied to MCMC-PF, can solve the tracking problem effectively, such as ; highly maneuvering target, track swapping. From the results obtained in the simulations, we have showed that in dense clutter environment, the MCMC-PF fails to tracks the high maneuvering targets and fails to avoid track swapping for the moving crossed target trajectory , where the proposed Viterbi filtered gate (VFG-MCMC) algorithm is capable of tracking the targets and avoid swapping. The VFG-MCMC avoids the false targets from the valid based measurement regions and chooses the best candidate valid measurement to the tracking filter MCMC-PF using VFG algorithm. Thus, this improves the data association process which has been shown to give targets the ability to continue to be tracked in dense clutter and also improves the tracking performance of MCMC-PF. This approach can be used to overcome the clutter of gate based approaches in tracking.

References

[1] Y. Bar-Shalom and T. E. Fortmann, “Tracking and data association”, academic press, 1988.

[2] Yaakov Bar-Shalom, Fred Daum, and Jim Huang. “The probabilistic

[3] data association filter", IEEE Control Systems Magazine Vol. 29, No. 6, PP. 82-100, December 2009.

[4] Blackman, Samuel, Robert Popoli, "Design and analysis of modern tracking systems", Boston: Artech House, 1999.

[5] Jaco Vermaak, Simon J. Godsill and Patrick P´erez, "Monte Carlo filtering for multi-target tracking and data association", IEEE Transactions On Aerospace And Electronic Systems Vol. 41, No. 1, 2005 , PP. 309-332

[6] G. W. Pulford "Multi-target Viterbi data association", Proc. of the 9th International Conference on Information Fusion, Florence, Italy, July 2006.

[7] E.M.Saad, El.Bardawiny, H.I.ALI and N.M.Shawky, " Improving data association based on finding optimum innovation applied to nearest neighbor for multi-target tracking in dense clutter environment", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No.1, PP.496-507, May 2011

[8] X. Wang, S. Challa, and R. Evans, "Gating techniques for maneuvering target tracking in clutter", IEEE Transactions on Aerospace and Electronic Systems, Vol..38, No.3, July 2002, PP. 1087–1097.

[9] G. David Forney, Jr. "The Viterbi algorithm". Proceedings of the IEEE, 61(3) , PP. 268–278, 1973.

[10] T. Quach and M. Farooq. "Maximum likelihood track formation with the Viterbi algorithm". In 33rd Conference on Decision and Control,CDC94, , Florida, USA, 1994, PP. 271–276.

[11] B. F. La Scala and G. W. Pulford. "Viterbi data association tracking for Over-The-Horizon Radar", In International Radar Symposium, IRS98, , Munich, Germany, 1998, PP. 1155–1164.

[12] Simon J. Julier and Je_rey K. Uhlmann. "A new extension of the Kalman filter to nonlinear systems", In Proc. AeroSense: 11th Int'l Symposium on Aerospace/Defense Sensing, Simulation and Controls, Orlando, Florida, 1997.

[13] E. A. Wan and R. van der Merwe, "The Unscented Kalman filter for nonlinear estimation", In Proceedings of IEEE Symposium 2000 on Adaptive Systems for Signal Processing, Communications and Control (AS-SPCC), Lake Louise, Canada, October 2000.

[14] Simon Haykin "Kalman filtering and neural networks" ,E-Book,Published by JOHN WILEY & SONS, INC.2001

[15] James V. Candy "Bayesian signal processing classical, modern, and particle filtering methods" ,E-Book, Published by JohnWiley & Sons, New Jersey 2009.

[16] B. Ristic, S. Arulampalam, and N. Gordon," Beyond the Kalman filter,particle filters for tracking applications", Boston, MA: Artech House, 2004.

[17] M. Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking" IEEE Transactions on Signal Processing, Vol. 50, No. 2, PP 174 -188, Feb.2002.

[18] C. Hue, J.-P. Le Cadre, P. Pérez. "A particle filter to track multiple objects". In IEEE Workshop on Multi-Object Tracking, PP. 61-68, Vancouver, Canada, July 2001.

[19] A. Doucet, N. de Freitas, and N. Gordon," Sequential Monte Carlo methods in practice", New York: Springer, 2001.

[20] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," Proc. Inst. Elect. Eng., Vol. 140, No. 2, pt. F, PP. 107–113, 1993.

[21] Jaward, M., Mihaylova, L., Canagarajah, N., & Bull, D. "Multiple object tracking using particle filters". IEEE Aerospace Conference, 1-8. Institute of Electrical and Electronics Engineers, 2006

[22] YANG Xiaojuna,*, XING Keyib, FENG Xinglea" Maneuvering target tracking in dense clutter based on particle filtering", Chinese Journal of Aeronautics Elsevier Vol.. 24, PP. 171-180, 2011.

[23] Zia Khan, Tucker Balch, Member, IEEE, and Frank Dellaert, Member, IEEE, "MCMC-based particle filtering for tracking a variable number of interacting targets" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.. 27, No. 11, PP. 1805-1819, Nov. 2005.

[24] Olivier Rabaste "Multi-target tracking with MCMC-based particle filters" 18th European Signal Processing Conference (EUSIPCO-2010) Aalborg, Denmark, August 23-27, 2010 , PP. 195-163.

[25] Hongtao Hu Zhongliang Jing Anping Li Shiqiang Hu Hongwei Tian "An MCMC-based particle filter for tracking target in glint noise environment " Proceedings The 7th International Conference on Information Fusion June 28 to July 1, 2004 in Stockholm, Sweden.

[26] Liu Jing, Prahlad Vadakkepat" Interacting MCMC particle filter for tracking maneuvering target" Elsevier Digital Signal Processing Vol. 20, No. 2, PP. 561–574, 2010.

[27] Franc¸ois Septier, Sze Kim Pang, Avishy Carmi and Simon Godsill "On MCMC-based particle methods for Bayesian filtering: application to multitarget tracking" 2009 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, , 2009, PP. 360-363.

[28] Gad.,Ahmed, and Farooq. M. "Single target tracking in clutter: performance comparison between PDA and VDA", proceeding of the sixth international conference of information fusion,Vol. .2 PP. 1266-1273, 2003.

[29] Ahmed Gad, F.Majdi M.Farooq "A comparison of data association techniques for target tracking in clutter", Proceedings of the Fifth IEEE International Conference on Information Fusion,Vol. 2, PP. 1126-1133, 2002.

[30] Y. Bar-Shalom and E. Tse, "Tracking in a cluttered environment with probabilistic date association",Automatica, Vol. 11, PP. 451-460, Sept. 1975.

Security Provisions in Stream Ciphers Through Self Shrinking and Alternating Step Generator

Hafsa Rafiq [#1], Malik Sikandar Hayat Kiyal[#2] , Aihab Khan [#3]
[#]Department of Software Engineering Fatima Jinnah Women University, The Mall, Rawalpindi, Pakistan

***Abstract*— in cryptography stream ciphers used to encrypt plain text data bits one by one. The security of stream ciphers depend upon randomness of key stream, good linear span and low probability of finding the initial states of pseudorandom generators. In this paper we propose a new stream cipher model use Feed back with carry shift registers (FCSRs) as building blocks which are considered as a source of long pseudorandom generator. Proposed model is the combined structure of alternating step generator and self shrinking generators which are commonly used stream ciphers generators. In this research we compare proposed generator against self shrinking generator (SSG), Alternating step generator (ASG) and alternating step self shrinking generator (ASSG) and we concludes that proposed generator architecture is best for encryption because it satisfies all necessary conditions of a good Stream cipher.**

Keywords-component; Alternating step generators; Feed back with carry shift registers; Self shrinking generators.

I. Introduction

Stream cipher is an important method for information encryption. “A stream cipher is a symmetric cipher which operates with a time-varying transformation on individual plaintext digits”[1]. Stream ciphers typically encrypt data efficiently and have very low memory requirements and therefore cheaper to implement in limited scenarios. Stream cipher techniques are usually best for the cases where the amount of data is either unknown, or continuous such as network streams.

Linear feedback shift registers (LFSRs) are mostly used in many key stream generators due to their simplicity but inherent linearity of LFSRs not sufficient to provide security to stream ciphers. Because of their linearity their initial vector can be determined by using Berlekamp-Massy algorithm. To improve the security a new type of pseudorandom binary sequence generator called FCSR is introduced. They have good statistical properties, having proved period, highly nonlinear in nature and have non degenerating states.[2].They are much similar to LFSRs except that instead of using addition modulo 2 FCSR uses carry propagations that bring non linearity in to their structure which is main characteristic of FCSR.

In this paper, we propose a model which combines two clock controlled keystream generators that are: alternating step generator and self shrinking generator to strengthen the security. Combined structure of both generators can help to avoid pitfalls that cause when these generators are used individually. Proposed generator is more secure because length of generated key stream is much greater than individual generators and complex structure is not easily breakable. Proposed model use FCSRs as a main building block and combine the output with full adder function.

The paper is organized in such a way that: section 2 discusses related work to proposed schemes, section 3 devoted to proposed framework, section 4 discuss proposed technique along with algorithm ,section 5 discuss simulation results and finally concluding remarks are given in section 6.

II. Related Work

Ali kenso [3] proposed a modified version of existing self shrinking generator proposed by Meier and Staffelbach based on selection rule which XOR pair of bits. MSSG constructed with single LFSR having length of n bits. The selection rule for the output of LFSR is that : select triple-bit($a_{3i,}a_{3i+1},a_{3i+2}$) then XOR first pair of bit a_{3i} and a_{3i} if result of XOR is 1 then output of MSSG becomes 3rd bit a_{3i+2} else discard triple bit .MSSG satisfies the basic requirements of good stream ciphers that are long period ,high complexity and non-linearity and proved that period p satisfies $p^{n/3}<=p<=p^{n-1}$ and linear complexity becomes greater than half the period. S.Shun-lung , Ko-ming Chiu, , and Lih-chyau Wuu [4], discuss “LFSR/FCSR based Alternating step generator” and a new combination function after analyzing existing ASG. According to the their analysis, the probability of finding the pairs of base sequence sequences (X_{n+1}, $_{Yn+1}$) which satisfy the condition of zero edit distance D(X_{n+1}, $_{Yn+1}$,K;Z_n) with exclusive-or operation is larger than addition operation. A new stream cipher generator is proposed by Ghosia Arshad [7], which is combination of shrinking generator and alternating step generator. Model is analyzed against co relational attacks and concluded that its security becomes 2^{2L}.

Model consists of 4 LFSRs A, B, C, D.LFSR A control sequence of LFSR B and C if output of LFSR A is 1 it activates LFSR B otherwise activate LFSR C. output of LFSR B and C is XORed if its result is 1 then generate output keystream of LFSR D else discard output of D.

III. Proposed Framework

The proposed model combines the features of both self shrinking generator proposed by Ali Kanso [3] and alternating step generator [4] to generate the strong pseudorandom keystream sequence. Proposed stream cipher system model based on FCSRs that introduce non linearity in proposed structure and make difficult to investigate the right initial states of registers due to carry propagations. To achieve the high security and to ensure important properties of FCSRs following conditions must be satisfied.

- Connection integer q must be negative prime number.
- Size of FCSR must be $r = \lfloor \log_2(q+1) \rfloor$ [5].
- Additional bits of memory must be $\lfloor \log_2 \lceil (r \rceil) \rfloor$
- The order of 2 modulo *q* is |*q*|-1[8].

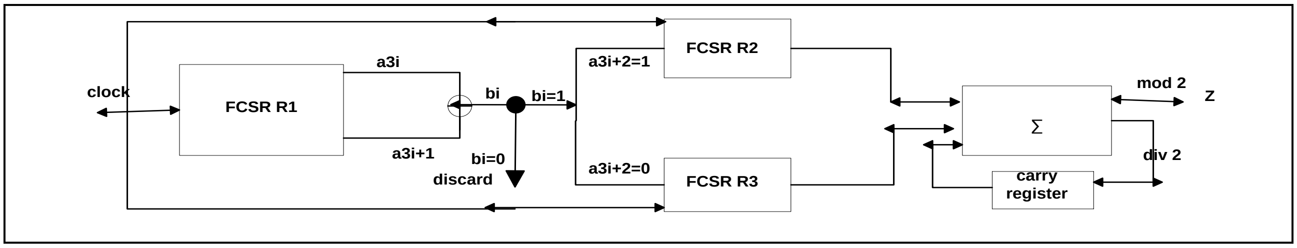


Figure 1: Detail structure of proposed model for stream ciphers generator

The Proposed model as shown in “Fig. 1”constructed in such a way that self shrinking generator is used to produce shrunken pseudorandom sequence and this shrunken output is used to control the remaining two registers. Output of these two registers is combined by full addition function and the resulting pseudorandom key stream is XORed with plain text to produce encrypted data. Resulting cipher data is again XORed with same key stream to obtain original plain text. Detail

IV. Proposed Technique

The graphical model of proposed generator is depicted in “Fig. 1”which shows that initial inputs to FCSRs are given, which generate sequences. The Generated sequence than produce key stream. Data from database is retrieved which is first converted to binary form and than encrypted with generated key stream.

A. Operational scenerio of proposed model

In this section we discuss the detail operation of our proposed generator .Initially before starting the process of key stream generation for each FCSR we should determine connection polynomial ***q***, size of register ***r*** and additional memory ***m.*** After that ***r*** cells of FCSRs are initialized with binary data.

1. Each FCSR process following steps:
   - Calculate $\sum_{k=1}^{\sigma n=r} q_k a_{n-k} + m_{n-1}$.
   - Output the rightmost bit a_{n-r} and all cells shift one step right.
   - Shift parity bit of σ_n (mod 2) in to leftmost cell and high order bit σ_n (div 2) in to carry cell.
2. Sequence of first FCSR is shrunken with self shrinking generator as follows:
   - Triple-bit ($a_{3i,}a_{3i+1},a_{3i+2}$) of first FCSR are taken at a time if ($a_{3i\ \mathrm{XOR}}\ a_{3i+1==1}$) than output becomes a_{3i+2} otherwise discard 3 bits .
3. Take shrunken sequence of FCSR as clock sequence and use to control other two base FCSRs sequences.
4. At each clock cycle output of two base sequences is added in to full addition function along with carry register. Output key stream is calculated as Z(mod2) and Carry register retained a new value Z (div2).
5. Output key stream is XORed with plain text bit to produce cipher text.
6. Plain text can be degenerate by XORing Cipher text with same keystream.

B. Algorithm:

Pseudo code of proposed model is

1) Algorithm: Input to connection number:

```
q=convert.Toint32 (q1text.Text)
FCSR_size =(int)(log2(q))
```

2) Algorithm: Input to FCSRs:

```
F1=R1.Text
```

3) Algorithm: Sequence Generation

```
SET seq=F1[0].ToString();
FOR i=initialR1.Length-1 TO 1
IF q[i] =1
SET sum+=F1[i];
END IF
ELSE
SET sum=0;
END ELSE
FOR temp=initialR1.Length-1 TO 1
F1 [temp]=F1[temp-1]
END FOR
```

```
        Sum=sum +m
        SET F1 [0] = mod (sum,2)
        m=div (sum, 2)
        END FOR
        DISPLAY seq;
```

4) // Algorithm: key Generation

a) SSG operation

```
        FOR i=0 TO SSG.Length-1
        b=a[3i]^a[3i+1]
        IF b=1
        Buffer[i]=a[3i+2]
```

b) ASG operation

```
        FOR j=0 TO buffer.Length-1
        DISPLAY Keystream
        IF buffer[j]=1
        Sum=R3prev+buffer[j]+carry
        END IF
        ELSE
        Sum=R2prev+buffer[j]+carry
        END ELSE
        Sum=mod(sum,2)
        Carry=div(sum,2)
        Keystream[i]=sum;
        END FOR
```

First input is given to connection integer which must be negative prime number then size of FCSR is calculated with connection integer. Initial value are given to each FCSR and with each FCSR we generate a unique sequence .

First FCSR behave like self shrinking generator and remaining two FCSRs act like alternating step generator. Output of these two FCSRs are combined with addition function.

V. Experiment Results and Analysis

A. Probability of finding initial states:

We can find the probability of success for finding correct initial values of registers R2 and R3 if addition modulo 2 is used as combined function as [4].

$$P = \frac{1}{2^{r2} + 2^{r3}} \qquad (1)$$

Here r2 and r3 are length of shift registers R1 and R2.

If full addition function is used as combination function then probability of success for initial states is determined as:

$$Padd = \frac{1}{2^{r2} + 2^{r3} + c1^{k1}} \qquad (2)$$

K1 is initial value of carry register of full adder function.

TABLE I. COMPARISON OF PROBABILITIES OF DIFFERENT GENERATORS

Generator	Probability
ASG	$Pasg = \frac{1}{2^{r2} + 2^{r3}}$
ASSG	$Passg = \frac{1}{2^{r2} + 2^{r3}}$
Proposed Generator	$Ppg = \frac{1}{2^{r2} + 2^{r3} + c1^{k1} + c1^{k2} + c1^{k3}}$

We have

$Ppg < pasg$ and $Ppg < Passg$

According to “Table 1” the structure of proposed generator is more secure than ASSG and SG because due to carry propagations it is difficult to estimate the right initial states of registers.

B. Graphical comparison ofprobabilty of proposed model with other models:

Case 1:

In case 1we compare the probabilities of success of finding right initial values of alternating step base registers.

For example by giving values 9 and 11 to ASG and ASSG registers r2 and r3 respectively we calculate probability as:

$$Pasg = Passg = \frac{1}{2^9 + 2^{11}} = 0.00039 \qquad (3)$$

Given values to proposed generator registers r2 r3 are 9 and 11.values to clock control sequence initial values of carry registers 1,2,3 are 7,1,2,3 respectively. then probability of Proposed generator(Ppg) becomes

$$Ppg = \frac{1}{2^9 + 2^{11} + 7^1 + 7^2 + 7^3} = 0.00037 \qquad (4)$$

So that Probability of finding right initial states is low in proposed model then other models as shown in “Fig 2” and its difficult to find initial states of registers .

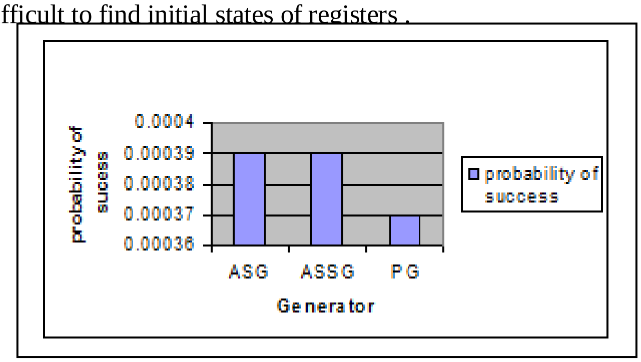


Figure 2: Case1: probability of success to find initial states

Case 2:

In case 2, we compare the probabilities of success of finding right initial values of alternating step base registers

For example, by giving values 3 and 5 to ASG and ASSG registers r2 and r3 respectively we calculate probability as:

$$Pasg = Passg = \frac{1}{2^3 + 2^5} = 0.025 \qquad (5)$$

Given values to proposed generator registers r2 r3 are 3 and 5.values to clock control sequence initial values of carry registers 1, 2, 3 are 5,1,2,0 respectively. Then probability becomes

$$Ppg = \frac{1}{2^3 + 2^5 \ + 5^0 + 5^1 + 5^2} = 0.0140 \qquad (6)$$

So that Probability of finding right initial states is low in proposed model then other models as shown in “Fig 3” ” and its difficult to find initial states of registers .

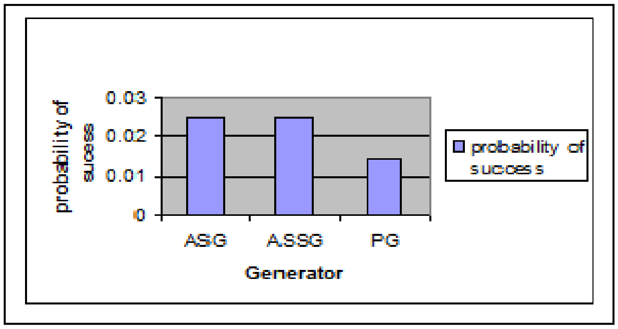


Figure 3: Case 2:probability of success to find initial states

C. *Comparison of proposed generator with other generators:*

Comparison of properties of proposed model with other models is shown in “Table 2”.

TABLE II. COMPARISON OF PROPOSED MODEL WITH OTHER MODELS

Properties	**SSG**	**ASG**	**SSG**	**Proposed generator**
Register type	LFSRs	LFSRs	LFSRs	FCSRs
number of registers	1	3	4	3
Logic gates	-	And ,Not	-	-
Combination function	-	XOR	XOR	Full adder, OR
structure	simple	complex	complex	complex
period	$(2^{L1-1}-1)$	$2^{L1}(2^{L2}-1)(2^{L3}-1)$	$2^{L1}(2^{L2}-1)(2^{L3}-1)(2^{L4}-1)$	$(\|q1\|-1)/3(\|q2\|-1)(\|q3\|-1)$

After comparing proposed generator with other generators it is found that proposed generator use FCSRs and full addition function which make structure highly non linear, so that proposed model is highly resistant to linear attacks as compared to other generators which uses LFSRs. Period of ASG and SSG is less than proposed generator while period of ASSG is greater because number of registers used in proposed generator less than ASSG registers which increase period of ASSG. Unless the period of proposed generator is less than ASSG it becomes more secure against attacks due to its non linearity [6].

VI. CONCLUSION

In this paper, we propose a model of keystream generator which is based on FCSRs and combine the features of both SSG and ASG to remove pitfalls that may occurs when these generators used individually. The description of proposed models and necessary conditions for model is well elaborated .The proposed model consists of 3 FCSRs which combine in such a way 1st FCSR used as clock sequence which controls the other 2 base FCSRs. Full addition function is used to combine the output of base registers.

Use of FCSRs, addition modulo2function and full addition function increase non linearity of proposed generator and make it more secure and highly resistant to external attacks .Next the comparison of proposed generator with other generators shows that it’s become a good choice because it is difficult to predict the right initial states of registers due to carry propagations in structure.

REFERENCES

[1] A. Rueppel, Analysis and Design of stream ciphers. Springer-Verlag,1986.

[2] Arnault, F., Berger, T.” F-FCSR: design of a new class of stream ciphers,” Lecture notes in computer sciences, vol. 3557, pp. 83–97. Springer, Heidelberg 2005.

[3] A.Kenso, “Modified self-shrinking generator,” Journal of Computers

[4] and Electrical Engineering vol 36, pp. 993–1001, 2010.

[5] S.Shun-lung , Ko-ming Chiu, , and Lih-chyau Wuu, “The Cryptanalysis of LFSR/FCSR Based Alternating Step Generator,” International conference on Computer Engineering and systems, pp. 228–231, 2006 .

[6] A. Klapper and M. Goresky,”Feedback shift registers, 2-adic span, and combiners with memory,”Journal of Cryptology, vol 10,pp.11–147, 1997.

[7] Lihua Dong, Yong Zeng, Yupu Hu, "F-GSS: A Novel FCSR-Based Keystream Generator," , First International Conference on Information Science and Engineering, pp.1737-1740, 2009.

[8] Ghosia Arshad “Analysis and design of alternating step shrinking generator ”,IJCTE,in press.

[9] Yong Zeng, Ma Jianfeng , Pei Qinqi, Dong Lihua “A Hardware-Oriented Fast Encryption Keystream Generator for Digital Rights Management”,International Conference on Computational Intelligence and Security,vol 02,pp. 584 – 586,2009

AUTHOR’S PROFILE

Hafsa Rafiq is a graduate from Dept. of Software Engineering, Fatima Jinnah Women University, Pakistan.

Dr. M. Sikandar H. Khiyal born at Khushab, Pakistan. He is Chairperson Dept. Computer Sciences and Software Engineering in Fatima Jinnah Women University Pakistan. He served in Pakistan Atomic Energy Commission for 24 years and involved in different research and development program of the PAEC. He developed software of underground flow and advanced fluid dynamic techniques. His area of interest is Numerical Analysis of Algorithm, Theory of Automata and Theory of Computation. He has more than one hundred research publications published in National and International Journals and Conference proceedings. He has supervised more than one hundred and thirty research projects at graduate and postgraduate level.

Mr.Aihab Khan works in Dept. of Computer Sciences at Fatima Jinnah Women University, Pakistan. His research interests are in the field of Data Mining, Data Warehousing as well as Information security.

Performance Analysis of Corporate Feed Rectangular Patch Element and Circular Patch Element 4x2 Microstrip Array Antennas

Md. Tanvir Ishtaique-ul Huque[.1], Md. Al-Amin Chowdhury [2], Md. Kamal Hosain[3] , Md. Shah Alam[4]
[1,2] Dept. of Electronics and Telecommunication Engineering,
[4] Dept. of Electrical and Electronic Engineering,
Rajshahi University of Engineering & Technology, Rajshahi-6204, Bangladesh.
[3] School of Engineering, Deakin University, Victoria-3217, Australia.

Abstract— **This paper present simple, slim, low cost and high gain circular patch and rectangular patch microstrip array antenna, with the details steps of design process, operate in X-band(8 GHz to 12 GHz) and it provides a mean to choose the effective one based on the performance analysis of both of these array antennas. The method of analysis, design and development of these array antennas are explained completely here and analyses are carried out for 4x2 arrays. The simulation has been performed by using commercially available antenna simulator, SONNET version V12.56, to compute the current distribution, return loss response and radiation pattern. The proposed antennas are designed by using Taconic TLY-5 dielectric substrate with permittivity, ε_r = 2.2 and height, h =1.588 mm. In all cases we get return losses in the range -4.96 dB to -25.21 dB at frequencies around 10 GHz. The gain of these antennas as simulated are found above 6 dB and side lobe label is maintained lower than main lobe. Operating frequency of these antennas is 10 GHz so these antennas are suitable for X-band application.**

Keywords*- microstrip array antenna; rectangular patch; return loss; X band; circular patch.***

I. Introduction

The term “Microstrip” comes because the thickness of this metallic strip is in micro-meter range. Microstrip patch antennas are popular, because they have some advantages due to their conformal and simple planar structure. They allow all the advantages of printed-circuit technology. A vast number of papers are available on the investigation of various aspects of microstrip antennas [1, 5, 6, 7, 8, 12, 13]. The key features of a microstrip antenna are relative ease of construction, light weight, low cost and either conformability to the mounting surface or, at least, an extremely thin protrusion from the surface. These criteria make it popular in the field of satellite and radar communication system. Different Radar systems such as synthetic aperture radar (SAR), remote sensing radars, shuttle imaging radar and other wireless communication systems operate in L, Ku, C and X bands [11, 12, 14, 15]. Microstrip antennas are the first choice for this high frequency band such as X-band due to its light weight, low cost, and robustness.

In this paper the designed microstrip antennas are also best suited for X band applications. The extended AM broadcast band or simply “X band” is a segment of the microwave radio region of the electromagnetic spectrum. X-band radar frequency sub-bands are used in civil, military and government institutions for weather monitoring, air traffic control, maritime vessel traffic control, defense tracking, and vehicle speed detection for law enforcement. In radar engineering, its frequency range is specified by the IEEE at 8.0 to 12.0 GHz. X band is used in radar applications including continuous-wave, pulsed, single-polarization, dual-polarization, synthetic aperture radar, and phased arrays. In Ireland, Libya, Saudi Arabia and Canada, the X band 10.15 to 10.7 segment is used for terrestrial broadband. Portions of the X band are assigned by the International Telecommunications Union (ITU) exclusively for deep space telecommunications. The primary user of this allocation is the American NASA Deep Space Network (DSN) [16].

Microstrip patch elements are available in various configuration. But the most common is the rectangular patch element and after the rectangular patch element the next most well known configuration is the circular patch element. This paper presents the design procedure, characteristic and the corresponding performance analysis of both the rectangular and circular patch microstrip array antennas and provides a mean to choose the effective one based on their performance parameters to get the efficient radiation efficiency.

In this paper we have also investigated the performance of corporate feed array in case of both the rectangular patch element and circular patch element, because it provides better directivity as well as radiation efficiency and reduce the beam fluctuations over a band of frequencies compared to the series feed array [5, 9]. Here all of these antennas are designed to support 10 GHz operating frequency and their corresponding simulations have been done by using the SONNET version V12.56 simulator. The proposed antennas are designed by using Taconic TLY-5 dielectric substrate with permittivity, ε_r = 2.2 and height, h =1.588 mm. These designed antennas are promising to be a good candidate for the X-band wireless applications due to the simplicity in structure, ease of fabrication and high gain and high efficiency.

II. MICROSTRIP ANTENNA DESIGN

Microstrip patch antennas consist of very thin metallic strip (patch) placed on ground plane where the thickness of the metallic strip is restricted by t<< λ_0 and the height is restricted by $0.0003\lambda_0 \leq h \leq .05\lambda_0$ [2, 5]. The microstrip patch is designed so that its radiation pattern maximum is normal to the patch. For a rectangular patch, the length L of the element is usually $\lambda_0 /3 < L < \lambda_0 /2$.

There are numerous substrates that can be used for the design of microstrip antennas and their dielectric constants are usually in the range of $2.2 \leq \varepsilon_r \leq 12$. To implement the microstrip antennas usually Fr-4 (ε_r= 4.9), Rogers TMM 4(ε_r= 4.5), Taconic TLY-5 (ε_r = 2.2), Alumina (96%) (ε_r = 9.4), Teflon(PTFE) (ε_r = 2.08), Arlon AD 5 (ε_r = 5.1) dielectric materials are used as the substrate [1, 2, 5, 10].

The Performance of the microstrip antenna depends on its dimension. Depending on the dimension the operating frequency, radiation efficiency, directivity, return loss and other related parameters are also influenced [3]. Here, in this paper, the investigation is made on two types of microstrip patch elements. They are

- Rectangular patch element
- Circular patch element.

A. Rectangular Patch Element

For an efficient radiation a practical width of the Rectangular patch element becomes [2, 3, 5]

$$w = \frac{1}{2 f_r \sqrt{\mu_o \varepsilon_o}} \times \sqrt{\frac{2}{\varepsilon_r + 1}} \qquad (1)$$

And the length of the antenna becomes [2, 3, 5]

$$L = \frac{1}{2 f_r \sqrt{\varepsilon_{eff}} \sqrt{\varepsilon_0 \mu_0}} - 2\Delta L \qquad (2)$$

Where,

$$\Delta L = 0.41h \frac{\varepsilon_{eff} + 0.3}{\varepsilon_{eff} - 0.258} * \frac{\left(\frac{w}{h} + 0.264\right)}{\left(\frac{w}{h} + 0.8\right)} \qquad (3)$$

And

$$\varepsilon_{eff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2\sqrt{1 + 12\frac{h}{w}}} \qquad (4)$$

Where, λ is the wave length, f_r (in Hz) is the resonant frequency, *L* and *W* are the length and width of the patch element, in cm, respectively and ε_r is the relative dielectric constant.

In the following Fig. 1, the antenna has been designed to cover specific 10 GHz operating frequency where the antenna dimension is in mm range and the quarter wavelength transformer method [2, 5] is used to match the impedance of the patch element with the transmission line.

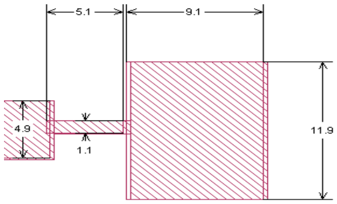


Figure 1. Single element Rectangular microstrip patch antenna.

B. Circular Patch Element

Other than the rectangular patch, the next most popular configuration is the circular patch or disk. For rectangular patch elements there are two degrees of freedom to control whereas for the circular patch elements there is one degree of freedom to control. Thus it is more convenient to design as well as to control the radiation pattern of the circular patch element.

From [12, 14] the first order approximation of the physical radius of the circular patch element becomes.

$$a = \frac{F}{\sqrt{\{1 + \frac{2h}{\pi \in_r F}[\ln(\frac{\pi F}{2h}) + 1.7726]\}}} \qquad (5)$$

Where
$$F = \frac{8.791 \times 10^9}{f_r \sqrt{\in_r}}$$

Thus the effective area of the circular patch element is given by [12]

$$A_{eff} = \pi a^2 \{1 + \frac{2h}{\pi \in_r F}[\ln(\frac{\pi F}{2h}) + 1.7726]\} \qquad (6)$$

Where, f_r (in Hz) is the resonant frequency, h(in cm) is the thickness of the substrate, ɑ is the effective radius of the circular patch element and ε_r is the relative dielectric constant.

In the following Fig. 2, the antenna has been designed to cover specific 10 GHz operating frequency where the antenna dimension is in mm range and the quarter wavelength transformer method [2, 5] is used to match the impedance of the patch element with the transmission line.

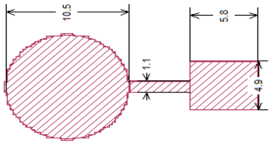


Figure 2. Single element circular microstrip patch antenna.

III. MICROSTRIP ARRAY ANTENNA DESIGN

Microstrip antennas are used not only as single element but also very popular in arrays. Microstrip arrays radiate efficiently only over a narrow band of frequencies and they can't operate at the high power levels of waveguide, coaxial line, or even stripline [1]. Antenna arrays are used to scan the beam of an antenna system, to increase the directivity and perform various other functions which would be difficult with any one single element. In the microstrip array, elements can be fed by a single line or by multiple lines in a feed network arrangement. Based on their feeding method [2, 5] the array is classified as

- Series feed network
- Corporate feed network

Series-feed microstrip array is formed by interconnecting all the elements with high impedance transmission line and feeding the power at the first element. Because the feed arrangement is compact the line losses associated with this type of array are lower than those of the corporate-feed type [5]. The main limitation in series-feed arrays is the large variation of the impedance and beam-pointing direction over a band of frequencies [5].

The corporate-feed network is used to provide power splits of 2*n* (i.e. *n* = 2*;* 4*;* 8*;* 16; etc.). This is accomplished by using either tapered lines or using quarter wavelength impedance transformers [5, 6]. In this paper the patch elements are connected by using the quarter wavelength impedance transformer method.

Corporate-feed arrays are general and versatile. This method has more control of the feed of each element and is ideal for scanning phased arrays, multiband arrays. Thus it provides better directivity as well as radiation efficiency and reduce the beam fluctuations over a band of frequencies compared to the series feed array [5, 9]. The phase of each element can be controlled by using phase shifters while amplitude can be adjusted using either amplifiers or attenuators [2, 8].

In this paper we have investigated the performance of corporate feed array in case of both the rectangular patch element and circular patch element.

A. *Rectangular Patch Microstrip Array*

In this paper we have designed the 8-elements rectangular patch Microstrip array antenna, as shown in Fig. 3, to cover 10 GHz operating frequency. Here the power is fed to the antenna by using the Microstrip transmission line method[2, 3] and the patch elements are matched together as well as with the transmission line with the quarter wavelength transformer method for the maximum power transmission.

The radiated field of the *E*-plane for a single element rectangular patch can be expressed by using the following formula [2, 10].

$$E = j\frac{k_0 W V_0 e^{-jk_0 r}}{r\pi}\{\frac{\sin(\frac{k_0 h}{2}\cos\varphi)}{\frac{k_0 h}{2}\cos\varphi}\}\cos(\frac{k_0 L_e}{2}\sin\varphi) \qquad (7)$$

Here, W is the width of the patch antenna, L_e is the extended length, V_0 =hE_0 is the voltage across radiating slot of the patch, h is the substrate height, K_0 =2π/λ and r is the far field distance from the antenna.

The array factor as given in [4, 8] as

$$FA = \frac{\sin^2(N\pi(d_x/\lambda)\sin\theta)}{N^2\sin^2(\pi(d_x/\lambda)\sin\theta)} \qquad (8)$$

Here, d_x is the element spacing and N is the number of elements. Combining array factor and element voltage radiation pattern we get the total element normalized power radiation pattern [2, 4, 5] that is

$$20\log(|E/FA|) \qquad (9)$$

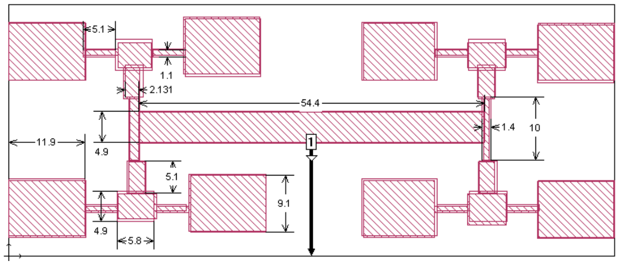


Figure 3. 8-elements corporate-feed rectangular microstrip array antenna.

B. *Circular Patch Microstrip Array*

Here the 8-elements circular patch Microstrip array antenna, as shown in Fig. 4, is designed to operate at 10 GHz frequency and similar to the previous one the power is fed to the antenna by using the Microstrip transmission line method and the patch elements are matched together as well as with the transmission line with the quarter wavelength transformer method for the maximum power transmission.

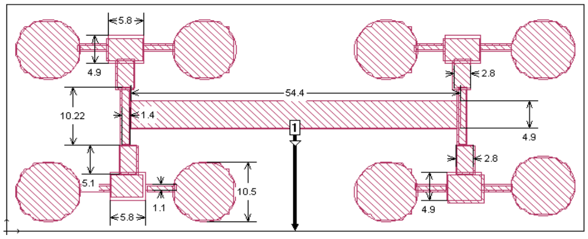


Figure 4. 8-elements corporate-feed circular microstrip array antenna.

The radiated field of the *E*-plane for a single element circular patch can be expressed by using the following formula [2, 5].

$$E = -jV_0\frac{ak_0 e^{-jk_0 r}}{2r}\cos\phi J_1^{'}(k_0 a\sin\theta) \qquad (10)$$

Here, θ is the beam pointing angle measured from the broadside direction, V_0 =hE_0 is the voltage across radiating slot of the patch, K_0 =2π/λ , J_1 is the Bessel function of first order and r is the far field distance from the antenna. The array factor as given in [14] as

$$AF = A_0\frac{\sin(\frac{N\psi}{2})}{N\sin(\frac{\psi}{2})} \qquad (11)$$

Where

$$\psi = \alpha + \beta d \sin(\theta)\cos(\varphi)$$

$$\beta d = (\frac{2\pi}{\lambda})S$$

Here, α is the phase difference between elements, N is the number of array elements and S is the spacing between circular patch elements. Now we can get the normalized power radiation pattern by combining the element radiation pattern and array factor [5].

IV. Simulation Result & Discussion

A. *Rectangular Microstrip Array Antenna Parameters*

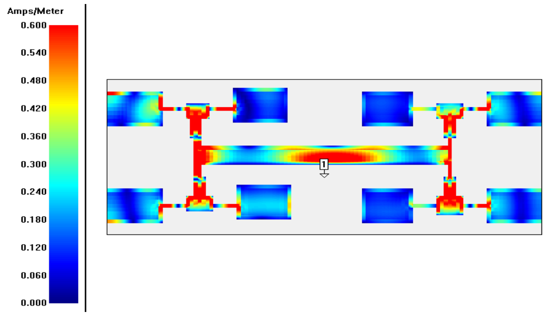


Figure 5. Current distribution of 8-elements rectangular microstrip array antenna.

In this paper, it is considered that the substrate permittivity of the antenna is ε_r = 2.2 (Taconic TLY-5), height is 1.588 mm and resonance frequency of the antenna is 10 GHz. After simulation, as shown in Fig. 6, we found that, return loss is -25.21 dB at 10 GHz and it is maximum. The simulated gain of the antenna, according to Fig. 7, is found around 6 dB at θ=0^0, φ=0^0 at the operating frequency 10 GHz. Fig. 8 shows that the HPBW and the FNBW for this simulated antenna are in the range 70^0& 142.10^0 respectively.

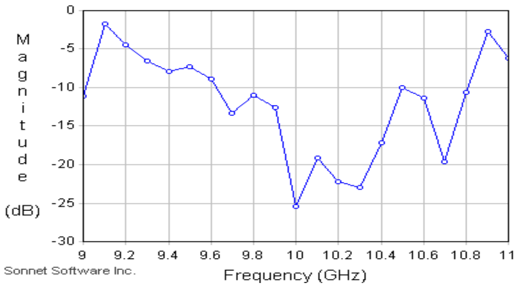


Figure 6. Return loss of the 8-elements rectangular microstrip array.

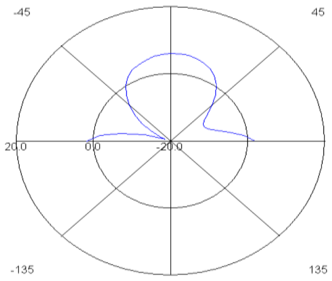


Figure 7. Radiation (Polar plot) pattern of the 8-elements rectangular microstrip array antenna.

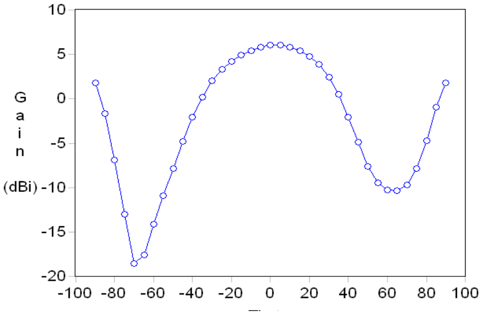


Figure 8. Radiation (Rectangular plot) pattern of the 8-elements rectangular microstrip array antenna.

B. *Circular Microstrip Array Antenna Parameters*

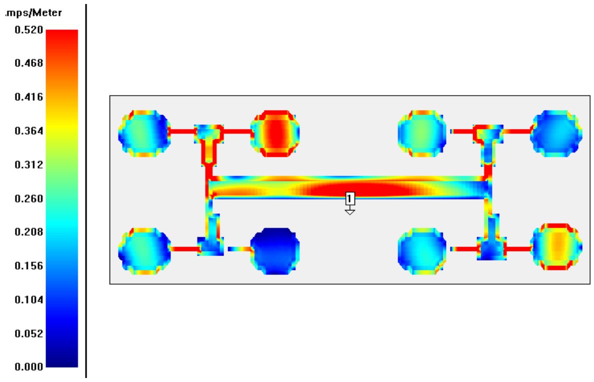


Figure 9. Current distribution of the 8-elements circular microstrip array antenna.

Here, the substrate permittivity of the antenna is 2.2 (Taconic TLY-5), height is 1.588 mm and resonance frequency of the antenna is 10 GHz. After simulation, as shown in Fig. 10, we found that, return loss is -4.96 dB at 10 GHz and it is maximum that is -6.91 dB at 9.8 GHz operating frequency. The simulated gain of the antenna, according to Fig. 11, is found around 7.21 dB at θ= 0^0 ,φ=0^0 for the operating frequency 10 GHz. Fig. 12 shows that the HPBW and the FNBW for this simulated antenna are in the range 89^0 and 130^0 respectively and Fig. 9 gives us the concept of current distribution of this array antenna whish states that current density at each element of circular array antenna is much higher than that of the rectangular one as shown in Fig. 5.

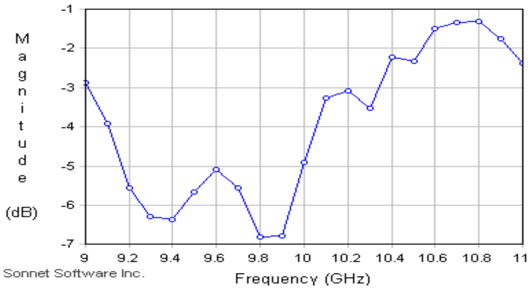


Figure 10. Return loss of the 8-elements circular microstrip array antenna.

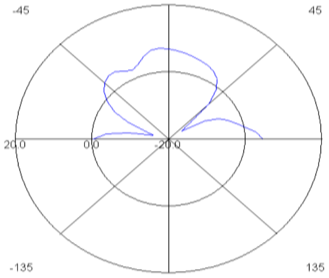


Figure 11. Radiation (polar plot) pattern of the 8-elements circular microstrip array antenna.

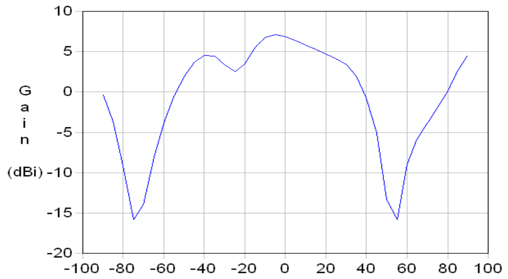


Figure 12. Radiation (rectangular plot) pattern of the 8-elements circular microstrip array antenna.

C. Comparison Between These two Array Antennas

After observing the performance analysis of both of these array antennas, it is convenient to say that the circular patch microstrip array antenna provides better performance than the rectangular patch Microstrip array antenna. Circular microstrip array antenna has the higher directive gain as well as the narrow beam width which seem to be a suitable criteria to design a transceiver antenna. Moreover its one degree of freedom to control reduce the design complexities. It also shows a remarkable achievements in case of current distribution. The current density for the rectangular microstrip array antenna, as shown in Fig. 5, at each patch element is not less than 0.12 *amp/m* and it is maximum up to 0.356 *amp/m.*

Whereas the current density for the circular microstrip array antenna, as shown in Fig. 9, at each patch element is not less than 0.156 *amp/m* and it is maximum up to 0.52 *amp/m*. Thus the circular Microstrip array occupies lower transmission loss and it's the higher current density at each element provides better radiation efficiency than the rectangular microstrip array.

TABLE I. PERFORMANCE ANALYSIS BETWEEN RECTANGULAR AND CIRCULAR MICROSTRIP ARRAY ANTENNAS

Performance Parameter	Feeding Types	
	Rectangular (4x2)array	Circular (4x2)array
Physical area(mm^2)	95.6×34.2	92.8×35.82
Degree of freedom to control	Two	One
FNBW	142.10^0	130^0
Simulated gain(dB)	Around 6	7.21
Transmission line loss	Higher	Lower
Return loss(dB)	-25.21	-4.96

V. CONCLUSION

The unique feature of this microstrip antenna is its simplicity to get higher performance. In many applications basically in radar and satellite communication, it is necessary to design antennas with very high directive characteristics to meet the demand of long distance communication and the most common configuration to satisfy this demand is the array form of the microstrip antenna.

After the rectangular patch the next most popular configuration is the circular patch and in our investigation, comparing the circular patch microstrip array antenna with the rectangular one, we have found that the circular microstrip array has some advantages such as small dimensions, light weight, higher directivity, higher current density and easy manufacturing. The physical area of the circular patch is 16% less than that of the rectangular patch [5].

Here designed array antennas covers 10 GHz operating frequency and it would also be possible to design the bands, operating any other system such as in WLAN, WiMax, WBAN or other wireless systems, by changing the dimension of the patch element. In future, we can investigate the spiral elements which seems to have more radiation efficiency for both the series feed and corporate feed networks and at the same time we can merge two different patch elements operating at two or more different frequencies by using quarter wavelength transformer method within an array network configuration to get multiband support.

REFERENCES

[1] R. J. Mailloux, J. F. Mcllvenna, N. P. Kernweis, "Microstrip array technology", IEEE Trans. Antenna Propagation Magazine, Vol. 29, No. 1, pp. 25-27, 1981.

[2] C. A. Balanis, Antenna Engineering, 2nd ed., Willey, 1982.

[3] T. A. Millikgan, Modern Antenna Design, 2nd ed. , IEEE Press, John Wiley & Sons inc., 2007.

[4] M. I. Skolnik, Introduction to RADAR System, 3rd ed., McGraw Hill Higher Education, 2000.

[5] R. Garg, P. Bhartia, I. Bahl, A. Ittipiboon, Microstrip Antenna Design Handbook, Artech House inc., 2001.

[6] R. J. Milloux, Electronically Scanned Arrays, Morgan & Claypool, 2007.

[7] W. L. Stutzman , "Estimating directivity and gain of antennas", IEEE Antennas and Propagation Magazine, Vol. 40, No. 4,pp 7-11, August, 1998.

[8] H. J. Visser, Array and Phased Array Antenna Basics, John Wiley & Sons Ltd., 2005.

[9] Muhammad Mahfuzul Alam, Md. Mustafizur Rahman Sonchoy, and Md. Osman Goni, "Design and Performance Analysis of Microstrip Array Antenna", Progress In Electromagnetic Research Symposium Proceedings, Moscow, Russia, August 18-21, 2009.

[10] Md. Shihabul Islam and Md. Tanvir Ishtaique-ul Huque, "Design and Performance Analysis of Microstrip Array Antenna", B.Sc. Engineering thesis, Dept. of ETE, Rajshahi University Of Engineering & Technology(RUET), Rajshahi, Bangladesh, April, 2010.

[11] Gi-cho Kang, Hak-young Lee, Jong-kyu Kim, Myun-joo Park,"Ku-band High Efficiency Antenna with Corporate-Series-Fed Microstrip Array", IEEE Antennas and Propagation Society International Symposium, 2003.

[12] T. F. Lai, Wan Nor Liza Mahadi, Norhayatision, "Circular Patch Microstrip Array Antenna for KU-band", World Academy of Science, Engineering and Technology, vol. 48, pp. 298-302, 2008.

[13] K. Shambavi, C. Z. Alex , T. N. P. Krishna, "Design and Analysis of High Gain Milimeter Wave Microstrip Antenna Array for Analysis of High Gain Millimeter Wave Microstrip Anteanna Array for Wireless Application", Journal of Applied Theoretical and Information Technology(JATIT), 2009.

[14] Asghar Keshtkar, Ahmed Keshtkar and A. R. Dastkhosh, "Circular Microstrip Patch Array Antenna for C-Band Altimeter System", International Journal of Antenna and Propagation, article ID 389418, doi:10.1155/2008/389418, November, 2007.

[15] M. F. Islam, M. A. Mohd. Ali, B. Y. Majlis and N. Misran, "Dual Band Microstrip Patch Antenna for Sar Applications", Australian Journal of Basic and Applied Sciences, 4(10): 4585- 4591, 2010.

[16] http://en.wikipedia.org/wiki/X_band

AUTHORS PROFILE

Md. Tanvir Ishtaique-ul Huque was born in 1988 in Bangladesh. He received his B.Sc. Engineering degree from the Rajshahi University of Engineering & Technology (RUET) in 2010. Now he is working as a part time teacher in the Dept. of Electronics and Telecommunication Engineering of RUET. His research interests include the antenna application of the wireless body area network(WBAN) and next generation wireless communication system.

Md. Al-Amin Chowdhury was born in 1988 in Bangladesh. he has completed his B.Sc. in Electronic and Telecommunicaton Engineering from Rajshahi University of Engineering & Technology(RUET) in 2010. He has keen interest to research on the optical fiber, different types of antennas. He wants to do his further study in USA on the communication field. World is becoming closer and closer due to the remarkable achievements in the communication field. He wants to receive the sound and proper knowledge in communication field so that he can contribute to the next generation demands in the communication sectors.

Md. kamal Hosain was born in 1982 in Bangladesh. He received his B.Sc. Engineering degree from the khulna University of Engineering & Technology(KUET) in 2001 and now he working as a Lecturer in the Dept. of Electronics and Telecommunication Engineering(ETE) of RUET. His research interests include the antenna and its application on the biomedical devices.

Md. Shah Alam was born in Rangpur, Bangladesh, on June 24, 1982. He received the B.Sc. degree in Electrical and Electronic Engineering from Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh, in 2006. He is going to be completed his M.SC degree from the same institution on July, 2011. From 2007 to 2009, he was a Lecturer with the Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh and currently he is working there as an Assistant Professor. His research interests involve Electromagnetic and Nanotechnology.

XML Based Representation of DFD

Removal of Diagramming Ambiguity

Swapna Salil Kolhatkar
Assistant Professor, Department of MCA, Marathwada Mitramandal's College Of Engineering,
Karvenagar,
Pune, India

***Abstract*—In the world of Information Technology, the working of a information system is well explained with the use of Data Flow Diagrams (DFD). DFDs are one of the three essential perspectives of the Structured Systems Analysis and Design Method (SSADM). [3]The sponsor of a project and the end users are briefed and consulted throughout all stages of a system's evolution. With a data flow diagram, users are able to visualize how the system will operate, what the system will accomplish, and how the system will be implemented. But, various practical problems exist with the representation of DFDs. Different tools are available in the market for representing the DFDs. These tools are user friendly and based on the object oriented features. The diagrams drawn using these tools can be sent over the network for communicating with others.**

On the other hand, the XML is platform independent, textual information and is totally extensible. XML is structured and is an excellent way to transfer the information along with the metadata over the network. XML can be used to transfer the information related to DFDs, there by removing the problems related to understanding the diagrammatic notations and concentrating more on the information flow in the system.

This paper is aimed at understanding the problems related to DFDs and representing it in XML[4] format for their storage and communication over the network. The discussion is divided into four main topics – introduction to XML and DFD, problems related to DFD, an XML representation for DFDs and finally the conclusion.

Keywords-data flow diagrams; XML; metadata; diagramming tools.

I. INTRODUCTION

The development and implementation of any successful IT system, requires communication and coordination amongst team members spread over different geographical locations across the world. SSADM is a well defined standard means of understanding the process and data flow in any system. The data flow diagrams are one of the three essential perspectives of SSADM. The DFDs are useful in visualizing the various processes in a system along with the data that flows within the system. The DFDs either follow the Gane and Sarson or DeMarco

and Yourdan method of representation which causes differences in their pictorial representation. In such a demanding scenario, many tools related to DFD that have been developed, may also differ. Users of such tools are used to working with pictorial representation that may differ depending upon the notations. Such diagrams have to be communicated over the network to other locations where the tools may differ.

Communication over the network has evolved over a time and in the current situation, XML has become a language of data communication over the network and also acts like a data store. Applications and systems communicate with each other using the XML format. Amongst these applications, those that are developed for diagramming purpose are already developed using the object oriented features. Such applications may store the data in XML format which is useful for network communication as well as for storage purpose. Let us consider a few important factors related to XML and DFD for our discussion related to the above.

- Extensible Markup Language (XML)

XML [1] is structured; XML documents are easily committed to a persistence layer, are platform independent, contain textual information, is an open standard, is language independent, totally extensible, supports share-able structure (using DTDs) and enables interoperability. The XML file (.xml) was designed to transport and store data. The Document Type Definition (DTD) helps to define the legal building blocks of an XML document along with the document structure with a list of legal elements and attributes. The schema document (XSD) is used to define the schema and data types for elements and attributes. Together, these three files provide the complete information on data and metadata. This paper focuses on these three main files for further discussion.

- Data Flow Diagrams (DFD)

The data flow diagram is a simple graphical technique which is easy to understand and helps in defining the boundaries of the system. It is useful for communicating current system knowledge to the users along with being a part of system documentation file so as to explain the logic behind the data flow within the system. There are different notations to draw data flow diagrams (Yourdon & Coad and Gane & Sarson – shown in fig 1.1f), defining different visual representations for processes, data stores, data flow, and external entities. These differences in notations have given rise to the problems [2] discussed in following section.

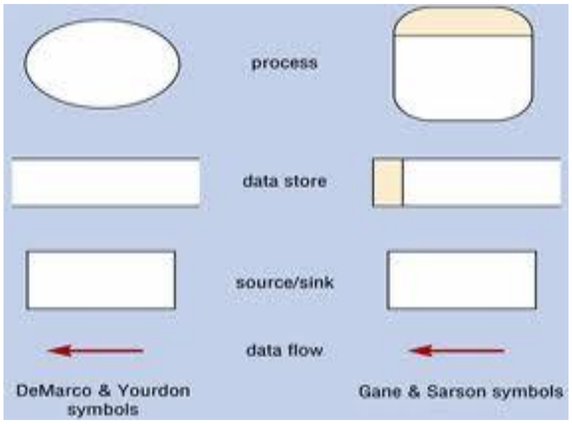


Fig 1.1f Differences in notations

II. Problems With DFD

- Time Consuming[2] - The Data flow diagrams undergo a lot of alteration before going to users, so as to make the process little slow. A faster and efficient communication over the network will surely help in reducing the delay.

- Ambiguity in understanding[2] - Different DFD models have different symbols like in Gane and Sarson process is represented as rectangle where as in DeMarco and Yourdan symbol it is represented as eclipse. These differences in pictorial representation cause ambiguity in system understanding at times.

III. XML Representation of DFDs.

Let us consider three XML based sample files : file1.xml, file2.dtd and file3.xsd

file1.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- New document created with EditiX at Thu
Aug 19 03:07:32 PDT 2010 -->
<!DOCTYPE mySystem SYSTEM
"PressAutoSystem.dtd">
<mySystem>
  <context>
     <process id="0">Press Automation
                            System</process>
        <entity id="S">Supplier</entity>
         ...............
       <dataflow id="df1">
          <process id="0"></process>
          <entity id="S"></entity>
          <direction>S</direction>
         <content>Request</content>
      </dataflow>
.......................
</context>
    <firstlevel>
       <process id="1">Registration</process>
          ..........................
       <entity id="S">Supplier</entity>
          ..........................
        <dataflow></dataflow>
     </firstlevel>
</mySystem>
```

The xml file ie file1.xml is actually used for sending the DFD related data or diagrams on the network. The file contains the metadata also in order to make the data self explanatory. The file can contain data related to context diagram (consisting of a single process, external entities and the data flows) and first level DFD (consisting of more detailed processes, external and internal entities, data stores). The same file can be used for the detailed logical and physical processes as well. As the level of detail goes on adding, the file consequently consists of other information.

file2.dtd

```
<!-- DTD created at Thu Aug 19 03:19:14 PDT 2010
with EditiX. Please insert an encoding attribute header
for converting any DTD  -->
<!ELEMENT  mySystem (context,firstlevel) >
<!ELEMENT context (process,entity+,dataflow+) >
<!ELEMENT process (#PCDATA) >
<!ATTLIST process id CDATA  "0" >
<!ELEMENT entity (#PCDATA) >
<!ATTLIST entity id CDATA "E1" >
<!ELEMENT dataflow
(process,entity,direction,content) >
<!ELEMENT direction (#PCDATA) >
<!ELEMENT content (#PCDATA) >
<!ATTLIST dataflow id CDATA "DF1" >
<!ELEMENT firstlevel
(process+,entity+,dataflow+,datastore+) >
<!ELEMENT datastore (#PCDATA) >
```

The second file ie file2.dtd is used for providing the rules for correctness of file1.xml. It specifies details related to the root element and its sub-elements along with details related to the number of occurrences of elements. This file is used in order to check the correctness of xml file at the client and server side.

The third file ie file3.xsd can be used additionally in order to specify the data types of the contents of elements. It can be also used to specify the optional values in the said diagrams.

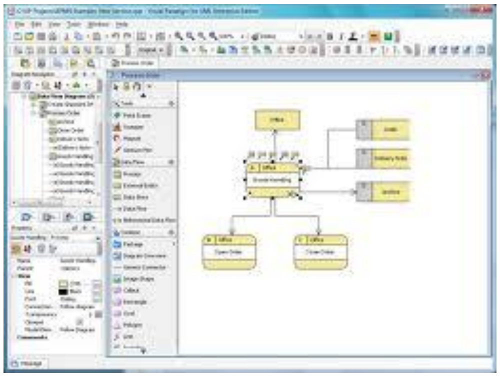


Fig 3.1 A Sample Tool for DFD

[5]Tools by which such diagrams can be drawn are any, for example Visio, Edraw, SmartDraw etc. A sample tool an be seen in Fig 3.1. Most of these tools are object oriented and are therefore easy to implement. Fig 3.2 shows a tool that saves data to an XML based file format. Accordingly a parser / translator that converts the data flow diagrams in XML format or reads an XML file for converting it into a diagram by identifying the diagramming notation that the user is acquainted with, can be provided additionally so as to take care of the problem of differences in diagramming notations.

Fig 3.2 EDraw Flowchart saves to an XML-based file format.

The communication over the network also becomes efficient due to the comparatively small size of xml files as compared to the size of files related to diagramming tools. XML files are easier to transport and improve readability of data due to the metadata.

IV. Conclusion

The current tools used for drawing of DFDs are quite user friendly and an addition of an interface to interpret the xml file for drawing of the diagrams would prove beneficial for speedy transportation as well as customizing the diagrammatic representation. The XML and its supporting formats together provide a correct data and metadata for parsing to be done at the client application. This removes any prerequisites of software tool understanding or implementation, thereby customizing the data flow diagrams as per the user's understanding.

XML technology is a very popular, efficient and effective technology for communication of the data over the network. The implementation of any diagrams as xml format will be very beneficial for handling such pictorial well defined data representation as data flow diagrams. It facilitates better system understanding by removing any ambiguities in the notations.

References

[1] w3schools – XML, DTD, XSD.

[2] Wiki answers.com : The_advantages and disadvantages of data flow diagrams

[3] Chris Gane and Trish Sarson. Structured Systems Analysis: Tools and Techniques. McDonnell Douglas Systems Integration Company, 1977

[4] The Relevance of the XML Data Format for Access to Historical Datasets and a Strategy for Digital Preservation : D-Lib Magazine, February 2005, Volume 11 Number 2, ISSN 1082-9873, The eXtensible Past

[5] Thinking XML: A glimpse into XML in the financial services industry Interesting lessons from a no-nonsense territory for XML Uche Ogbuji (uche@ogbuji.net), Principal Consultant, Fourthought, Inc.

[6] An introduction to Model Driven Architecture Part I: MDA and today's systems Alan Brown (dwinfo@us.ibm.com), Staff, IBM

http://www.ibm.com/developerworks/rational/library/3100.html

[7] The Importance of Data Modeling as a Foundation for Business Insight by Larissa Moss, Method Focus, Inc. and Steve Hoberman, Steve Hoberman & Associates, LLC

Some Modification in ID-Based Public key Cryptosystem using IFP and DDLP

Chandrashekhar Meshram
Department of Applied Mathematics
Shri Shankaracharya Engineering College, Junwani,
Bhilai (C.G.), India

S.A. Meshram
Department of Mathematics
R.T.M.Nagpur University,
Nagpur (M.H.) India

***Abstract*— In 1984, Shamir [1] introduced the concept of an identity-based cryptosystem. In this system, each user needs to visit a key authentication center (KAC) and identify him self before joining a communication network. Once a user is accepted, the KAC will provide him with a secret key. In this way, if a user wants to communicate with others, he only needs to know the "identity" of his communication partner and the public key of the KAC. There is no public file required in this system. However, Shamir did not succeed in constructing an identity based cryptosystem, but only in constructing an identity-based signature scheme. Meshram and Agrawal [5] have proposed an id - based cryptosystem based on integer factoring and double discrete logarithm problem which uses the public key cryptosystem based on integer factoring and double discrete logarithm problem. In this paper, we propose the modification in an id based cryptosystem based on the integer factoring and double discrete logarithm problem and we consider the security against a conspiracy of some entities in the proposed system and show the possibility of establishing a more secure system.**

Keywords- Public key Cryptosystem; Identity based Cryptosystem; Discrete Logarithm Problem (DLP); Double Discrete Logarithm Problem (DDLP); Integer Factorization Problem (IFP).

I. INTRODUCTION

In an open network environment, secret session key needs to be shared between two users before it establishes a secret communication. While the number of users in the network is increasing, key distribution will become a serious problem. In 1976, Diffie and Hellman [6] introduced the concept of the public key distribution system (PKDS). In the PKDS, each user needs to select a secret key and compute a corresponding public key and store in the public directory. The common secrete session key, which will be shared between two users can then be determined by either user, based on his own secret key and the partner's public key. Although the PKDS provides an elegant way to solve the key distribution problem, the major concern is the authentication of the public keys used in the cryptographic algorithm.

Many attempts have been made to deal with the public key authentication issue. Kohnfelder [7] used the RSA digital signature scheme to provide public key certification. His system involves two kinds of public key cryptography: one is in modulo p, where p is a large prime number; the other is in modulo n, where n = p q, and p and q are large primes. Blom [11] proposed a symmetric key generation system (SKGS based on secret sharing schemes. The problems of SKGS however, are the difficulty of choosing a suitable threshold value and the requirement of large memory space for storing the secret shadow of each user.

In 1984, Shamir [1] introduced the concept of an identity. In this system, each user needs to visit a key authentication center (KAC) and identify himself before joining the network. Once a user's identity is accepted, the KAC will provide him with a secret key. In this way, a user needs only to know the "identity" of his communication partner and the public key of the KAC, together with his secret key, to communicate with others. There is no public file required in this system. However, Shamir did not succeed in constructing an identity-based cryptosystem, but only in constructing an identity-based signature scheme. Since then, much research has been devoted, especially in Japan, to various kinds of ID-based cryptographic schemes. Okamoto et al. [10] proposed an identity-based key distribution system in 1988, and later, Ohta [12] extended their scheme for user identification. These schemes use the RSA public key cryptosystem [18] for operations in modular n, where n is a product of two large primes, and the security of these schemes is based on the computational difficulty of factoring this large composite number n. Tsujii and Itoh [2] have proposed an ID- based cryptosystem based on the discrete logarithm problem with single discrete exponent which uses the ElGamal public key cryptosystem.

In 2004, Wei Bin lee & Kuan Chieh Liao [13] design a transformation process that can transfer all of the discrete logarithm based cryptosystems into the ID-based systems rather than reinvent a new system .After 2004 Several ID-Based cryptosystems [21,22,23, 24, 25, 26] have been proposed. But in these schemes, the public key of each entity is not only an identity, but also some random number selected either by the entity or by the trusted authority. Meshram [27] have also proposed Cryptosystem based on double generalized discrete logarithm problem whose security is based on double generalized discrete logarithm problem with distinct discrete exponents in the multiplicative group of finite fields. Meshram & Agrawal [4] have also proposed an ID- based cryptosystem based on the double discrete logarithm problem with double distinct discrete exponent which uses the Public key cryptosystem based on the double discrete logarithm problem.

Meshram and Agrawal [5] have proposed an id - based cryptosystem based on integer factoring and double discrete logarithm problem which uses the public key cryptosystem based on integer factoring and double discrete logarithm problem. Now we Modified this cryptosystem for integer factoring and discrete logarithm problem with distinct double discrete exponent because we face the problem of solving integer factoring and discrete logarithm problem simultaneously in the multiplicative group of finite fields as compared to the other public key cryptosystem. Where, we face the difficulty of solving the traditional discrete logarithm problem in the common group.

In this Study, we present modification in an ID based cryptosystem based on the integer factoring and double discrete logarithm with distinct discrete exponent (the basic idea of the proposed system comes on the public key cryptosystem based on integer factoring and double discrete logarithm problem) here we describe further considerations such as the security of the system, the identification for senders. etc. our scheme does not require any interactive preliminary communications in each message transmission and any assumption except the intractability of the discrete logarithm problem and integer factoring (this assumption seems to be quite reasonable)Thus the proposed scheme is a concrete example of an ID –based cryptosystem which satisfies Shamir's original concept [1] in a strict sense.

II. Modified ID-Based Public key Cryptosystem

A. Implementation of the ID –Based Cryptosystem

1) Preparation for the center and each entity

Step 1: Each entity generates a k-dimensional binary vector for his ID. We denote entity A's ID by ID_A as follows

$$ID_A = \left(x_{A1}, x_{A2}, \ldots\ldots\ldots, x_{Ak}\right), x_{Aj} \in \{0,1\}, \ (1 \le j \le k) \qquad (1)$$

Each entity registers his ID with the center, and the center stores it in a public file.

Step 2: The center generates two random prime numbers p and q and compute

$$N = pq \qquad (2)$$

Then the center chooses an arbitrary random number $e, 1 \le e \le \phi(N)$, such that $\gcd\left(e, \phi\left(N\right)\right) = 1$ where $\phi(N) = (p-1)(q-1)$ is the Euler function of N. Then center publishes (e, N) as the public key. Any entity can compute the entity $A's$ extended ID, EID_A by the following:

$$EID_A \equiv (ID)^e (\bmod N)$$

$$= \left(y_{A1}, y_{A2}, \ldots\ldots\ldots, y_{At}\right), x_{Aj} \in \{0,1\} \ , \ (1 \le j \le t)$$

(3)

where $t = |N|$ is the numbers of bits of N.

Step3: *Center's secrete information:* - The center chooses an arbitrary large prime number p and q and compute $N = pq$ and also generated n-dimensional vector a and m-dimensional vector b over $z^*_{\phi(N)}$ which satisfies

$$a = \left(a_1, a_2, \ldots\ldots\ldots\ldots, a_n\right), b = \left(b_1, b_2, \ldots\ldots\ldots\ldots, b_m\right) \qquad (4)$$

$$2 \le a_i b_l \le \phi(N) - 1 \ _, (1 \le i \le n), (1 \le l \le m), (m \le n)$$

$$abI \ne abJ (\bmod (p-1)), I \ne J \qquad (5)$$

Where I and J are n-dimensional binary vector and stores it as the centers secret information. The condition of equation (5) is necessary to avoid the accidental coincidence of some entities secrete key. A simple ways to generate the vectors a and b is to use Merkle and Hellmans scheme [19].

Step 4: The center also chooses w which satisfies $\gcd(w, \varphi(N)) = 1$ and $w < \lfloor \varphi(N)/n \rfloor$, where $\lfloor x \rfloor$ also denote the floor function which implies the largest integer smaller than compute x.

The center chooses a super increasing sequences corresponding to a and b as $a_i^{'} (1 \le i \le n)$ & $b_l^{'} (1 \le l \le m)$ satisfies

$$\sum_{j=1}^{i-1, l-1} a_j^{'} b_j^{'} + v \ \text{p} \ \varphi(N) \quad \text{,where } v = \lfloor \varphi(N)/w \rfloor \qquad (6)$$

$$\sum_{j=1}^{n} a_j^{'} b_j^{'} \ \text{p} \ \varphi(N), (m \le n) \qquad (7)$$

Then the centre computes

$$aibl = a_i^{'} b_l^{'} w (\bmod \varphi(N))$$

$$c_i = a_i b_l (\bmod w)(1 \le i \le n)(1 \le l \le m)(m \le n) \qquad (8)$$

Where

$$a = \left(a_1, a_2, \ldots\ldots\ldots\ldots, a_n\right) \ b = \left(b_1, b_2, \ldots\ldots\ldots\ldots, b_m\right) \qquad (9)$$

Remark 1: it is clear that the vector a and b defined by (9) satisfies (4)-(5) the above scheme is one method of generating n and m dimensional vectors a and b satisfies (4)-(5). In this paper, we adopt the above scheme. However, another method might be possible.

***Step 5*:** The center also chooses an unique integer d d, $1 \le d \le \varphi(N)$ such that $ed \equiv 1 (\bmod \varphi(N))$ and arbitrary integer t such that $e = \left(e_1, e_2, \ldots\ldots\ldots\ldots, e_t\right)$, satisfying $\gcd(e_i, \varphi(N)) = 1, \ (1 \le i \le t)$ and compute n-dimensional and m- dimensional vectors D^j and D^k respectively:

$$D^j = (d_1^j, d_2^j, \ldots\ldots d_n^j)(1 \le j \le n)$$

$$d_l^j = e_l a_l (\bmod \varphi(N))(1 \le l \le n) \qquad (10)$$

$$D^k = (d_1^k, d_2^k, \ldots\ldots d_m^k)(1 \le k \le m)$$

$$d_l^k = e_l b_l (\bmod \varphi(N))(1 \le l \le m)(m \le n) \qquad (11)$$

Since D^j and D^k are one to one system.

Step 6: *Center public information:* The center chooses two arbitrary generators α and β of $Z^*_{\varphi(N)}$ and computes n-dimensional vector h using generator α & m-dimensional vector g using generator β corresponding to the vector a and b.

$$h = \left(h_1, h_2, \ldots\ldots\ldots\ldots, h_n\right), g = \left(g_1, g_2, \ldots\ldots\ldots\ldots, g_m\right) \qquad (12)$$

$$h_i = \alpha^{a_i} (\bmod N), (1 \le i \le n),$$

$$g_l = \beta^{b_l} (\bmod N), (1 \le l \le m) \qquad (13)$$

The center informs each entity $\left(N, \alpha, \beta, h, g\right)$ as public information.

Step 7 Each entity secrete key: Entity $A's$ secrete keys s_a and s_b are given by inner product of a and b (the centre's secret information) and EID_A (entity $A's$ extended ID, see eqn.3)

$$s_a \equiv d_l^j EID_A (\bmod \phi(N))$$
$$= \sum_{1 \le j \le n} d_l^j y_{Aj} (\bmod \phi(N))$$

(14)

$$s_b \equiv d_l^k EID_A (\bmod \phi(N))$$
$$= \sum_{1 \le j \le n} d_l^k y_{Aj} (\bmod \phi(N))$$

(15)

2) *System Initialization Parameters*
Center Secrete information

a : n -dimensional vector and b m-dimensional vector and d an integer {see (8)-(9)}

Center public information

h : n -dimensional vector & g m-dimensional vector {see eqn.(12-13)} p and q :large prime numbers, e : random integers , two generator α and β of $z^*_{\phi(N)}$.

Entity $A's$ secrete keys s_a and s_b = entity $A's$ public information = ID_A, k-dimensional vector

III. Protocol of the Proposed Cryptosystem

Without loss of generality supposes that entity B wishes to send message M to entity A.

A. *Encryption*

Entity B generates EID_A (Entity $A's$ extended ID, see eqn.3) from ID_A . It then computes γ_1 and γ_2 from corresponding public information h and g and EID_A .

$$\gamma_1 = \Big(\prod_{1 \le i \le n} h_i^{\,y_{Ai}}\Big)^{ei} (\bmod N)$$
$$= \Big(\prod_{1 \le i \le n} (\alpha^{a_i})^{y_{Ai}}\Big)^{ei} (\bmod N)$$
$$= \alpha^{\sum_{1 \le i \le n} ei\alpha i y_{Ai} (\bmod \varphi(N))} (\bmod N)$$
$$= \alpha^{\sum_{1 \le i \le n} d_i^j y_{Ai} (\bmod \varphi(N))} (\bmod N)$$
$$= \alpha^{sa} (\bmod N)$$

$$\gamma_2 = \Big(\prod_{1 \le l \le m} g_l^{\,y_{Al}}\Big)^{el} (\bmod N)$$
$$= \Big(\prod_{1 \le l \le m} (\beta^{b_l})^{y_{Al}}\Big)^{el} (\bmod N)$$
$$= \beta^{\sum_{1 \le l \le m} el\beta l y_{Al} (\bmod \varphi(N))} (\bmod N)$$
$$= \beta^{\sum_{1 \le l \le m} d_l^k y_{Al} (\bmod \varphi(N))} (\bmod N)$$
$$= \beta^{sb} (\bmod N)$$

Entity B use γ_1 and γ_2 in Public key cryptosystem based on double discrete logarithm problem.

Let $M(1 \le M \le N)$ be entity B's message to be transmitted. Entity B select two random integer u and v such that $(2 \le uv \le \varphi(N) - 1)$ and computes

$$Y1 = \alpha^u (\bmod N)$$
$$Y2 = \beta^v (\bmod N)$$
$$\delta = M(\gamma_1)^u (\gamma_2)^v (\bmod N)$$

$$= M(Y_1^{sa} Y_2^{sb})(\bmod N)$$

And compute

$$C_1 = (Y_1)^e(\bmod N)$$

$$C_2 = (Y_2)^e(\bmod N)$$

$$E = (\delta)^e(\bmod N)$$

The cipher text is given by $C = (C_1, C_2, E)$.

B. Decryption

To recover the plaintext M from the cipher text

Entity A should do the following Compute

$$C_1^{\phi(N)-s_a}(\bmod N) = C_1^{-s_a}(\bmod N)$$

And $C_2^{\phi(N)-s_b}(\bmod N) = C_2^{-s_b}(\bmod N)$

Recover the plaintext $M = \left(C_1^{-s_a} C_2^{-s_b} E\right)^d (\bmod N)$

IV. Security Analysis

In this section, we shall show three possible attacks by which an adversary may try to take down the new encryption scheme. For each attack, we define the attack and give reason why this attack could be failed.

A. Direct Attack

Adversary wishes to obtain all secrete keys using all information available from the system. In this case, adversary needs to solve factoring and discrete logarithm problem with double distinct discrete exponent. The best way to factorize is by using the number field sieve method (NFS) [28].but this method is just dependent on the size of modulus n.It is computationally infeasible to factor a 1024-bit integer and to increase the security of our scheme; we should select strong primes [29] to avid attacks using special purpose factorization algorithms. To maintain the same security level for discrete logarithm problem with double distinct discrete exponent, one must uses with and respectively is product of two 512-bit primes.

B. Factoring Attack

Assume that the adversary successfully solves the factoring problem so that he knows secrete d. Thus he may obtain

$$\left(C_1^{-s_a} C_2^{-s_b} E\right)^d (\bmod N) = M^{ed}(\bmod N)$$

Unfortunately, at this stage he still does not knows secrete a and b and cannot extract the plaintext M from the above expression.

C. Discrete log Attack

An attacker should solve a discrete logarithm problem twice to obtain the private key given the public as following:

1) An attacker should solve a discrete logarithm problem twice to obtain the private key given the public as following:

In this encryption the public key is given by $\left(N, e, \alpha, \beta, \gamma_1, \gamma_2\right)$ and the corresponding secret key is given by $\left(s_a, s_b\right)$.

To obtain the private key $\left(s_a\right)$ he should solve the DLP

$$s_a \equiv \log_\alpha\left(\alpha^{s_a}\right)(\bmod N)$$

To obtain the private key $\left(s_b\right)$ he should solve the DLP

$$s_b \equiv \log_\beta\left(\beta^{s_b}\right)(\bmod N)$$

This information is equivalent to computing the discrete logarithm problem over multiplicative cyclic group $z_{\phi(N)}^*$ and corresponding secrete key s_a and s_b will never be revealed to the public.

2) An attacker might try to impersonate user A by developing some relation between w and w'

since $\gamma_1 \equiv Y^{ws_a}(\bmod N)$ and $\gamma_1' \equiv Y^{w's_a}(\bmod N)$

Similarly $\gamma_2 \equiv Y^{ws_b}(\bmod N)$ and $\gamma_2' \equiv Y^{w's_b}(\bmod N)$ by knowing $\gamma_1, \gamma_2, w, w'$ the intruder can derive γ_1' and $'\gamma_2'$ as $\gamma_1' = \gamma_1^{w^{-1}w'}(\bmod N)$ and $\gamma_2' = \gamma_2^{w^{-1}w'}(\bmod N)$ without knowing s_a and s_b however trying to obtain w from α and β is equivalent to compute the discrete logarithm problem.

V. Conclusion

In this study, some modification in an ID-based cryptosystem based on integer factoring and double discrete logarithm problem with distinct discrete exponents in the multiplicative group of finite fields. The proposed scheme satisfies Shamir's original concepts in a strict sense, i.e. it does not require any interactive preliminary communications in each data transmission and has no assumption that tamper free modules are available. This kind of scheme definitely provides a new scheme with a longer and higher level of security than that based on integer factoring and double discrete logarithm problem with distinct discrete exponents.

The proposed scheme also requires minimal operations in encryption and decryption algorithms and thus makes it is very efficient. The present paper provides the special result from the security point of view, because we face the problem of solving integer factoring and double and triple distinct discrete logarithm problem simultaneously in the multiplicative group of finite fields as compared to the other public key cryptosystem.

REFERENCES

[1] A. Shamir "Identity-based cryptosystem and signature scheme," Advances in Cryptology: Proceedings of Crypto' (Lecture Notes in Computer Science 196). Berlin, West Germany: Springer-Verlag, vol. 84 pp. 47-53,1985.

[2] S. Tsujii, and T. Itoh "An ID-Based Cryptosystem based on the Discrete Logarithm Problem"IEEE Jounral on selected areas in communications vol. 7 pp 467-473, 1989.

[3] T. ElGmal "A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms", IEEE Trans. Inform. Theory, vol. 31, pp 469-472, 1995

[4] C.S.Meshram and S.S.Agrawal "An ID-Based Public key Cryptosystem based on the Double Discrete Logarithm Problem" International Journal of Computer Science and Network Security, vol.10 (7) pp.8-13,2010.

[5] C.S.Meshram and S.S.Agrawal "An ID-Based Public key Cryptosystem based on Integer Factoring and Double Discrete Logarithm Problem" Information Assurance and Security Letters, vol.1 pp.029-034,2010.

[6] W. Diffie and M.E. Hellman, "New direction in Cryptography", IEEE Trans.Inform.Theory, vol. 22, pp 644-654,1976.

[7] L. M. Kohnfelder, "A method for certification," Lab. Comput. Sci. Mass. Inst. Technol.. Cambridge, MA, May 1978.

[8] S. Tsujii, T. Itoh, and K. Kurosawa, "ID-based cryptosystem using discrete logarithm problem," Electron. Lett., vol. 23. no. 24, pp 1318-1320,1987.

[9] S. C. Pohlig and M. E. Hellman, "An improved algorithm for com puting logarithms over GF (p) and its cryptographic significance," IEEE Trans. Inform. Theory, vol. IT-24, pp. 106-110,1978.

[10] E. Okarnoto and K. Tanaka, "Key distribution system based on identification information," IEEE J. SeIecr. Areas Commun., vol. 7, pp.481485, May 1989.

[11] R. Blorn, "An optimal class of symmetric key generation systems." In Proc. Eurocryp '84, Pans, France, Apr. 9-11, pp. 335-338,1984.

[12] K. Ohta, "Efficient identification and signature schemes." Electron. Lett., vol. 24, no. 2, pp. 115-116,1988.

[13] Wei-Bin Lee and Kuan-Chieh Liao "Constructing identity-based cryptosystems for discrete logarithm based cryptosystems" Journal of Network and Computer Applications,vol. 27, pp. 191–199,2004.

[14] Min-Shiang Hwang, Jung-Wen Lo and Shu-Chen Lin "An efficient user identification scheme based on ID-based cryptosystem" Computer Standards & Interfaces,vol. 26,pp. 565–569,2004.

[15] Eun-Kyung Ryu and Kee-Young Yoo "On the security of efficient user identification scheme" Applied Mathematics and Computation 2005, vol.171, pp. 1201–1205.

[16] Mihir Bellare , Chanathip Namprempre and Gregory Neven "Security Proofs for Identity-Based Identification and Signature Schemes" J. Cryptol.,vol. 22, pp. 1–61, 2009.

[17] S. C. Pohlig and M. E. Hellman, "An improved algorithm for computing logarithms over GF (p) and its cryptographic significance," IEEE Trans. Inform. Theory, vol. IT-24, pp. 106-110,1978.

[18] R. L. Rivest, A. Shamir And L. Adelman, "A method for obtaining digital signatures and public-key cryptosystem," Comrnun. ACM., vol. 21, no. 2, pp. 120-126,1978.

[19] R. C. Merkle and M. E. Hellman, "Hiding information and signatures in trapdoor knapsacks" IEEE Trans. Inform. Theory, vol. IT- 24, pp. 525-530,1978.

[20] C.S.Laih and J.Y.Lee "Modified ID-Based Public key Cryptosystem using Discrete Logarithm Problem" Electronic Letters, vol.24 (14) pp.858-859,1988.

[21] Min-Shiang Hwang, Jung-Wen Lo and Shu-Chen Lin "An efficient user identification scheme based on ID-based cryptosystem" Computer Standards & Interfaces, vol. 26, pp. 565–569, 2004.

[22] Eike Kiltz and Yevgeniy Vahlis. "CCA2 Secure IBE: Standard model efficiency through authenticated symmetric encryption" In CT-RSA, Vol. 4964 of Lecture Notes in Computer Science, pp 221–239. Springer,2008.

[23] Raju Gangishetti, M. Choudary Gorantla, Manik Lal Das, Ashutosh Saxena "Threshold key issuing in identity-based cryptosystems" Computer Standards & Interfaces, vol.29, pp.260–264,2007.

[24] Jinyuan Sun, Chi Zhang, Yanchao Zhang, and Yuguang Fang"An Identity-Based Security System for User Privacy in Vehicular Ad Hoc Networks" IEEE Tran. On Parall. and Distributed Systems, vol.27, no.9,pp. 1227-1239,2010.

[25] Dan Boneh and Matthew K. Franklin. "Identity based encryption from the Weil pairing" SIAM Journal on Computing, Vol.32 (3), pp.586–615,2003.

[26] Dan Boneh, Ran Canetti, Shai Halevi, and Jonathan Katz "Chosen-ciphertext security from identity-based encryption" SIAM Journal on Computing,Vol.5(36), pp.1301–1328,2006.

[27] Chandrashekhar Meshram "A Cryptosystem based on Double Generalized Discrete Logarithm Problem" Int. J. Contemp. Math. Sciences,Vol. 6, no. 6, 285 -297,2011.

AUTHORS PROFILE

Chandrashekhar Meshram received the M.Sc and M.Phil degrees, from Pandit Ravishankar Shukla University, Raipur (C.G.) in 2007 and 2008, respectively and pursuing PhD from R.T.M. Nagpur University, Nagpur (M.H.) India. Presently he is teaching as an Assistant Professor in Department of Applied Mathematics, Shri Shankaracharya Engineering College, Junwani Bhilai, (C.G.) India. He is doing his research in the field of Cryptography and its Application. He is a member of International Association of Engineers, Hong Kong, Computer Science Teachers Association (CSTA)USA, Association for Computing Machinery (ACM) USA ,International Association of Computer Science and Information Technology (IACSIT), Singapore, European Association for Theoretical Computer Science (EATCS) Greece, International Association of Railway Operations Research (IAROR) Netherland, International Association for Pattern Recognition (IAPR) New York and International Federation for Information Processing (IFIP) Austria, International Mathematical Union , International Linear Algebra Society (ILAS) and Life -time member of Internet Society (ISOC) USA ,Indian Mathematical Society ,Cryptology Research Society of India and Ramanujan Mathematical Society of India (RMS). He is regular reviewer of ten International Journals and International Conferences.

Dr. (Mrs) S. A. Meshram received the MSc, M.Phil and PhD degrees, from R.T.M. Nagpur University, Nagpur (M.H.) India. Presently she is teaching as Associate Professor in Department of Mathematics, R.T.M. Nagpur University, Nagpur (M.H.) and is having 27 years of teaching experience postgraduate level in University. She is carrying out her research work in the field of Thermo elasticity, Solid Mechanics, Cryptography and its Application. Dr. Meshram published eighteen research papers in National and International Journals.

Simulation of Spectral Subtraction Based Noise Reduction Method

Zhixin Chen
ILX Lightwave Corporation
Bozeman, Montana, USA
chen.zhixin.mt@gmail.com

***Abstract* — Noise reduction is a very meaningful but difficult task and it has been a subject of intense research in recent years. This paper introduces two popular noise reduction techniques and presents our simulation result of a noise reduction system. It is shown that the system reduces the noise almost completely while keeps the enhanced speech signal very similar to the original speech signal.**

Keywords - Embedded Systems; Digital Signal Processing; Noise Reduction.

I. INTRODUCTION

Noise is one kind of sound that is unexpected or undesired [1, 10]. It can be divided into non-additive noise and additive noise. The non-additive noise includes multiplier noise and convolutional noise, which can be transformed into additive noise through homomorphic transform. The additive noise includes periodical noise, pulse noise, and broadband noise [2, 4]. The noise generated by the engine is one kind of periodical noise while the one generated from explosion, bump, or discharge is pulse noise. There are many kinds of broadband noise, which may include heat noise, wind noise, quantization noise, and all kinds of random noise such as white noise and pink noise.

Loud and persistent noise can be very dangerous to human’s health since it influences the function of the human body system and thus causes the body to respond in ways that may lead to stress disorder, irritability, and sleepiness [5]. Noise with sufficient intensity can temporally or even permanently damage hearing.

Consequently, noise has gradually become a new dangerous pollutant. Fortunately, people start to notice this and they have taken an action against excessive noise in recent years [7].

The ideal way to treat the noise pollution would be just to get rid of it. This can be done for environmental noise like traffic noise or machinery noise by using a barrier material or a passive absorber [1]. However, in most cases, a noise polluted signal is more complicated because it is desirable to cancel the noise completely but retain the desired signal. An example is the detection of a weak signal in a noisy environment or a telephone interview in the noisy environments such as cars, airports and laboratories [8]. We can imagine the sad expression of the interviewee if we design a system to get rid of all the sound, which includes the noise and his conversation with the interviewer. The noise reduction is a very meaningful but difficult task and it has been a subject of intense research in recent years.

This paper introduces two popular noise reduction techniques and presents our simulation result of a noise reduction system.

The remaining part of this paper is organized as follows. First, the popular noise reduction techniques, including spectral subtraction and Wiener filtering, are reviewed. Then, the simulation of a spectral subtraction based noise reduction system is presented. Finally, this paper is concluded with a summary of results.

II. NOISE REDUCTION TECHNIQUES

In a phone interview in a noisy environment, the speech signal may be distorted by the background noise, which can be generated by the machine, computer, or even the electronic fans. If a handsfree telephone is used in the interview, the intensity of the background noise may be even stronger than the speech signal. The noise will thus distort the speech and make it hardly intelligible. In order to improve the intelligibility, the noise needs to be attenuated to enhance the speech signal. Figure 1 shows the block diagram of a noise reduction system. In this figure, the noisy speech signal X(n) is the combination of the original speech signal S(n) and the noise N(n). The noisy speech signal X(n) passes through a noise reduction system to get a clean speech signal Y(n), which is similar to the original speech signal S(n).

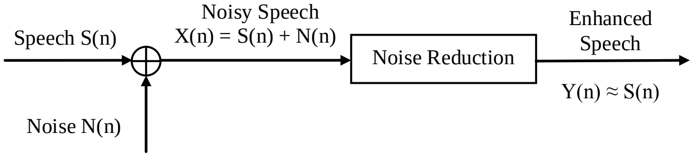


Figure 1: The block diagram for noise reduction system

The intelligibility and naturalness of the enhanced signal, improvement of signal-to-noise ratio, signal delay, and computational complexity are four key criteria for the performance of a noise reduction system [9]. It is obvious that after noise reduction, a high quality speech with high intelligibility and naturalness is desired. At the same time, the interference needs to be as weak as possible, which also means that a high signal-to-noise ratio is needed. It is reported that if the delay for the telephone communication is larger than 100 milliseconds, the delay can be perceived [3, 6]. Consequently, short signal delay and low computational complexity are also desired in the noise reduction systems.

Noise reduction algorithms generally operate in the frequency domain, which includes adaptive filtration, autocorrelation methods, methods based on parametric models of spectrum estimation, and techniques employing intelligent algorithms [9, 11]. The existing noise reduction systems are mainly based on spectral subtraction and Wiener filtering.

A. *Spectral Subtraction*

Spectral subtraction is the most popular noise reduction method [2, 9, 11]. This method operates in the frequency domain and assumes that the spectrum of the input noisy signal can be expressed as the sum of the speech spectrum and the noise spectrum. Figure 2 shows the block diagram for the spectral subtraction method. The noise spectrum is first estimated and then subtracted from the noisy speech spectrum to get the clean speech spectrum.

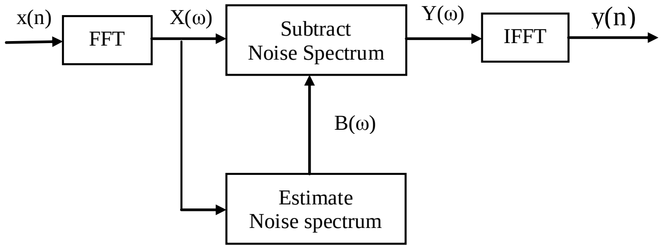


Figure 2: the block diagram for the spectral subtraction method

In Figure 2, it is assumed that x(n) is a discrete-time noisy sequence with x(n) = s(n) + b(n), where s(n) is the desired signal and b(n) is the unwanted background noise. Here s(n) and b(n) are assumed to be wise-sense stationary and uncorrelated random processes with power spectral density functions denoted by $S_s(\omega)$ and $S_b(\omega)$, respectively. The sum of the power spectra $S_x(\omega) = S_s(\omega) + S_b(\omega)$ is used to recover the desired signal s(n). Although speech signal is time-variant but it can be considered stationary in a short time interval. The short time Fourier transform (STFT) analysis can be used in the short-time segments. In the time domain, $x_{pL}(n) = w[pL - n](s(n) + b(n))$, where L is frame length and p is an integer. The frequency domain expression is $X(pL,\omega) = S(pL,\omega) + B(pL,\omega)$, where $S(pL,\omega)$, $B(pL,\omega)$ and $X(pL,\omega)$ are the STFTs of the object s(n), the background noise b(n), and the measurement x(n) computed at frame interval L, respectively. So, the STFT magnitude squared of x(n) can be written as:

$$|X(pL,\omega)|^2 = |S(pL,\omega)|^2 + |B(pL,\omega)|^2 + S^*(pL,\omega)B(pL,\omega) + S(pL,\omega)B^*(pL,\omega)$$
(1)

Here the objective is to obtain an estimate of $|S(pL,\omega)|^2$.

In this signal estimation approach, the STFT phase is not estimated. Consequently, the best for each short-time segment is an estimate of the form $\hat{S}(pL,\omega) = |S(pL,\omega)| e^{j\angle X(pL,\omega)}$. This means that the ideal STFT estimate consists of the clean STFT magnitude and noisy measure STFT phase. This is referred as the theoretical limit in estimating the original STFT when only the STFT magnitude is estimated. By considering the threshold of perception of phase deviation due to additive noise, it has been shown that speech degradation is not perceived with an average short-time segmental SNR greater than 6dB for the theoretical limit in the equation above-mentioned. However, when this SNR is considerably below 6 dB, a roughness of the reconstruction will be perceived.

Assume that an estimate of the power spectrum of the noise, denoted by $\hat{S}_b(\omega)$, which is typically obtained by averaging over multiple frames of a known noise segment, is given. Also assume that the noise and object sequences are uncorrelated. Then with short-time analysis, an estimate of the object's short-time squared spectral magnitude is expressed as

$$|\hat{S}(pL,\omega)|^2 = \begin{cases} |X(pL,\omega)|^2 - \hat{S}_b(\omega) & \textit{if} \quad |Y(pL,\omega)|^2 - \hat{S}_b(\omega) \geq 0 \\ 0 & \textit{otherwise} \end{cases}$$
(2)

When this magnitude estimate is combined with the measured

phase, the STFT estimate $\hat{S}(pL,\omega) = |\hat{S}(pL,\omega)| e^{j\angle X(pL,\omega)}$ is obtained. An objected signal estimate can then be formed with overlap-add (OLA), filter-bank summation (FBS), or least-squared-error (LSE) synthesis [2, 11].

B. *Wiener Filtering*

An alternative way to spectral substraction for recovering an object suquence s(n) from a sequence x(n) = s(n) + b(n) is to find a linear filter h(n) such that the sequence $\hat{s}(n) = x(n) * h(n)$ minimizes the expected value of $|s(n) - \hat{s}(n)|^2$ [7]. Under the condition that the signals s(n) and b(n) are uncorrelated and stationary, the frequency-domain solution to this stochastic optimization problem is given by the suppression filter

$$H_s(\omega) = \frac{S_s(\omega)}{S_s(\omega) + S_b(\omega)}, \qquad (3)$$

which is referred to as the Wiener filter [9, 11].

When the signals s(n) and b(n) are uncorrelated and stationary, the Wiener filter provides noise suppression without considerable distortion in the estimated object. The required power spectra, $S_s(\omega)$ and $S_b(\omega)$, can be estimated by averaging over multiple frames when sample functions of s(n) and b(n) are provided. But the desired signal and background are typically nonstationary in the sense that their power spectra change over time, which also means that they can be expressed as time-varying functions $S_s(n,\omega)$ and $S_b(n,\omega)$. Consequently, each frame of the STFT is processed by a different Wiener filter. For the simplified case of a stationary background, the time-varying Wiener Filtering can be expressed as:

$$H_s(pL,\omega)=\frac{\hat{S}_s(pL,\omega)}{\hat{S}_s(pL,\omega)+\hat{S}_b(\omega)},$$
(4)

where $\hat{S}_s(pL,\omega)$ is an estimate of $S_s(n,\omega)$ on each frame and $\hat{S}_b(\omega)$ is an estimate of $S_b(\omega)$. As with spectral subtraction, for the Wiener filtering, an enhanced waveform is recovered from the modified STFT, $\hat{S}(pL,\omega)=Y(pL,\omega)H_s(pL,\omega)$, by overlap-add (OLA), filter-bank summation (FBS), or least-squared-error (LSE) synthesis.

III. Simulation Of Noise Reduction

One thing for both spectral subtraction and Wiener filtering is how to estimate the noise. According to [9, 11], there are two main methods. The simpler form is the analysis during speech pauses while the more complicated one is a minimum statistics algorithm. Another thing for spectral subtraction and Wiener filtering is that their processed output signals suffer from musical noise. These artifacts are due to randomly distributed spectral peaks in the residual noise spectrum. In [12], Ephraim and Malah proposed two popular methods to reduce the musical noise phenomenon.

In this section, the simulation result of a noise reduction system is given. This system is based on spectral subtraction. It processes the noisy speech signal with a sampling frequency of 8 KHz or 16 KHz. It first transforms the signal from time domain to frequency domain and then divides the frequency domain into 16 or 19 channels, depending on sampling frequency. Then it estimates the noise signal during speech pause and subtracts the noise from the noisy speech in every channel. Finally, it sums the signal from every channel and then converts the signal from the time domain back to frequency domain. This algorithm performs well in modest noisy environment with low computational complexity. The subjective reconstructed speech quality is good. Figure 3 lists the waveform and spectrogram of the noisy speech and reconstructed speech. It can be seen from both the waveform and the spectrogram that this system reduces the noise almost completely while keeps a speech signal very similar to original speech signal.

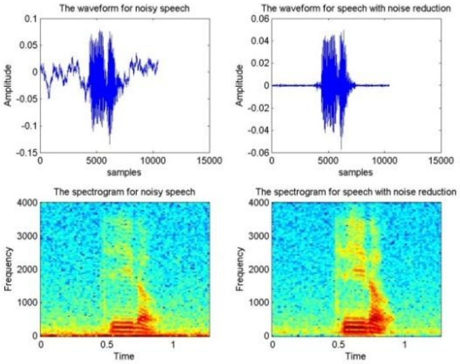


Figure 3: The waveform and spectrogram of the noisy speech and enhanced speech.

IV. Conclusion

This paper introduced the definition, classification, and the hazard of noise. It then described two main techniques, spectral subtraction and Wiener filtering, for noise reduction. The paper also presented the simulation result for a noise reduction system based on spectral subtraction.

Currently, there is a lot of research on the new techniques for noise reduction. First, the spectral subtraction and Wiener filtering did not rely on a speech model. Researchers found that a noise reduction filter that exploits estimated speech model parameters can be designed in the noise reduction system.

For example, the Wiener filter can be constructed with an object power spectrum estimation that is based on an all-pole vocal tract transfer function. This filter can then be applied to enhance speech, just like the nonparametic case.

Second, in the phenomenon of auditory masking, one sound component is concealed by the presence of another sound component. Hence, the auditory masking principle can be used in reducing the perception of noise.

Finally, the wavelet transform can also be used to replace the STFT method for noise reduction.

References

[1] Acoustical Solutions, Inc. “Introduction to Noise Control”, Acoustic Education.

[2] Yang Xingjun, Chi Huisheng, “Digital Speech Signal Processing”, 1992 (Chinese).

[3] Zhixin Chen, "Design and implementation on a sub-band based acoustic echo cancellation approach," International Journal of Advanced Computer Science and Applications, vol. 2, issue 6, 2011.

[4] Kinsler, Lawrence etc., “Fundamentals of Acoustics”, 4th ed., Wiley & Sons, 1999.

[5] The Kansas City Health Department, “Noise Pollution”.

[6] Zhixin Chen, "Investigation on simulation and measurement of reverberation for small rooms," International Journal of Advanced Computer Science and Applications, vol. 2, issue 7, 2011.

[7] “Introduction to Noise Control”, http://www.polytechinc.com/noise.htm

[8] Joachim Holzfuss, “Active Noise Reduction”, 1997.

[9] Stefan Schmitt, Malte Sandrock, “Single Channel Noise Reduction Algorithm for Handsfree Operation in Distorted Enviroments”, 2001.

[10] Zhixin Chen and Robert C. Maher, "Analytical expression for impulse response between two nodes in 2-D rectangular digital waveguide mesh," IEEE Signal Processing Letters, vol. 15, pp. 221-224, 2008.

[11] Thomas F. Quatieri, “Discrete-Time Speech Signal Processing”, Prentice Hall Signal Processing Seriers.

[12] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator", IEEE Transactions on Acoustics, Speech, and Signal processing, vol. 32, no. 6, pp. 1109-1121, December 1984

Authors Profile

Zhixin Chen is a Firmware Engineer in ILX Lightwave Corporation. He holds a BS degree from Xiamen University in China, a MS degree from Xiamen University in China, and a Ph.D. degree from Montana State University in USA, all in Electrical and Computer Engineering.

His research interest and working experience are in the area of acoustics, audio, and speech processing, multimedia communication, and embedded system design for high power current source, temperature controller, and optical power meter.

Implementation Of Node Energy Based On Encryption Keying

Dr.S.Bhargavi
Electronics and Communication Engineering
S.J.C.I.T
Chikballapur, Karnataka, India

Ranjitha B.T
Electronics and Communication Engineering
S.J.C.I.T
Chikballapur, Karnataka, India

***Abstract*—This paper deals with Designing cost-efficient, secure network protocols for any Networks is a challenging problem because node in a network itself is resource-limited. Since the communication cost is the most dominant factor in any network, we introduce an energy-efficient Node Energy-Based Encryption and Keying (NEBEK) scheme that significantly reduces the number of transmissions needed for rekeying to avoid stale keys. NEBEK is a secure communication framework where sensed data is encoded using a scheme based on a permutation code generated via the RC4 encryption mechanism. The key to the RC4 encryption mechanism dynamically changes as a function of the residual energy of the node. Thus, a one-time dynamic key is employed for one packet only and different keys are used for the successive packets of the stream. The intermediate nodes along the path to the sink are able to verify the authenticity and integrity of the incoming packets using a predicted value of the key generated by the sender's virtual energy, thus requiring no need for specific reeking messages. NEBEK is able to efficiently detect and filter false data injected into the network by malicious outsiders. We have evaluated NEBEK's feasibility and performance analytically and through software simulations. Our results show that NEBEK, without incurring transmission overhead (increasing packet size or sending control messages for rekeying), is able to eliminate malicious data from the network in an energy efficient manner.**

Keywords- NEBEK; Network; protocol; communication; RC4 encryption; dynamic key; virtual energy; Statistical mode; Operational mode; Forwarding Node Packets.

I. INTRODUCTION

From a security point of view, it is important to provide authentic and accurate data to surrounding nodes and to the sink. Protocols should be such that they are resilient against false data injected into the network by malicious nodes. Else the consequences of propagating a false data in a network become costly, depleting the network resources and wasting responses. This becomes a challenging to the protocol builder in securing the network.

Here we focus on 2 keying mechanisms. Static and Dynamic keying. In static scheme keys are handling statistically. i.e. the network node will have fixed no of keys loaded. But dynamic after key revocation. Thus refreshed key doesn't become any stale key. Here we focus on minimizing the overhead associated with refreshing keys since the communication cost is the most dominant factor. This scheme performs keying function either periodically or on demand needed by the network. The major drawback of this keying mechanism is that it increases the communication overhead due to keys being refreshed in a network. Key refreshment may require for updating key.

In this project we develop an efficient and secure communication framework for network. Here we introduce NEBEK for network.

II. LITERATURE SURVEY

A. Problem Statement

Sending confidential information from one node (source) to another node (destination) on a network could be a challenging task. Using the available resources and energy, the nodes exchange data of the received and sent packets and also ensure data integrity before it hits the sink.

The data exchanged could be manipulated or changed by the hacker on the network. So, the task would be to create a secure system that can ensure safety of the data using encryption methods (such as RC4) and still use the available energy and resources without much overhead.

B. Objective of the Paper

The objective of this paper is to discuss efficient and secure communication frameworks for Network applications by building upon the idea of sharing a dynamic cryptic credential.

Designing cost-efficient, secure network protocols for any Networks is a challenging problem because all the networks are resource-limited. Since the communication cost is the most dominant factor in a energy consumption, it is necessary to introduce an energy-efficient Node Energy-Based Encryption and Keying (NEBEK) scheme for LAN network that significantly reduces the number of transmissions needed for rekeying to avoid stale keys.

C. Existing System

An existing Dynamic Energy-based Encoding and Filtering (DEEF) framework is to detect the injection of false data into a sensor network. Dynamic Energy-based that each sensed event report be encoded using a simple encoding scheme based on a keyed hash.

The key to the hashing function dynamically changes as a function of the transient energy of the nodes, thus requiring no need for re-keying. Depending on the cost of transmission vs.

computational cost of encoding, it may be important to remove data as quickly as possible. Accordingly, DEEF can provide authentication at the edge of the network. Depending on the optimal configuration, as the report is forwarded, each node along the way verifies the correctness of the encoding probabilistically and drops those that are invalid.

Disadvantages

- Current schemes involve the usage of authentication keys and secret keys to disseminate the authentication keys; hence, it uses many keys and is complicated for resource-limited nodes.
- Current schemes are complicated for resource-constrained sensors as they transmit many keying messages in the network, which increases the energy consumption of WSNs that are already severely limited in the technical capabilities and resources (i.e., power, computational capacities, and memory) available to them.

D. *Proposed System*

NEBEK is a secure communication framework where the data is encoded using a scheme based on a permutation code generated via the RC4 encryption mechanism. The key to the RC4 encryption mechanism dynamically changes as a function of the residual energy of the network. Thus, a one-time dynamic key is employed for one packet only and different keys are used for the successive packets of the stream.

The intermediate nodes along the path to the sink are able to verify the authenticity and integrity of the incoming packets using a predicted value of the key generated by the sender's virtual energy, thus requiring no need for specific rekeying messages.

NEBEK's flexible architecture allows for adoption of stronger encryption mechanisms in lieu of encoding. And also show that our framework performs better than other comparable schemes in the literature with an overall 60-100 percent improvement in energy savings without the assumption of a reliable medium access control layer.

Advantages

- NEBEK's secure communication framework provides a technique to verify data in line and drop false packets from malicious nodes, thus maintaining the health of the wireless network.
- NEBEK dynamically updates keys without exchanging messages for key renewals and embeds integrity into packets as opposed to enlarging the packet by appending message authentication codes (MACs).
- The key to the encryption scheme (RC4) dynamically changes as a function of the residual virtual energy of the node, thus requiring no need for rekeying.
- The protocol is able to continue its operations under dire communication cases as it may be operating in a high-error-prone deployment area like under water.

III. System Description

A. *Node Energy-Based Keying Module*

NEBEK is a simple idea of designing the secure communication framework. It provides a technique to verify data in line and drop false packets from the malicious node, thus maintaining the security of network. Here data is encoded using RC4 encryption mechanism. RC4 mechanism dynamically changes as a function of residual energy of the network. The Node energy-based keying process involves the creation of dynamic keys. Here, it does not exchange extra messages to establish keys unlike other dynamic scheme methodologies. A node computes keys based on its residual energy of the network [5].

The rationale for using node energy as opposed to real battery levels as in our earlier work, DEEF [4], is that in reality battery levels may fluctuate and the differences in battery levels across nodes may spur synchronization problems, which can cause packet drops. These concerns have been addressed in NEBEK. After deployment, each nodes traverse several functional states. The states mainly include node-stay-alive, packet reception, transmission, encoding, and decoding. As each of these actions occur, the energy in a node is depleted. The current value of the node energy, Evc, in the node is used as the key to the key generation function, F. During the initial deployment, each node in a network will have the same energy level Eini, therefore, the initial key, K1, is a function of the initial virtual energy value and an initialization vector (IV).The IVs are pre distributed to the all the nodes. Subsequent keys, Kj, are a function of the current virtual energy, Evc, and the previous key Kj_1.

B. *Operation mode of NEBEK*

The NEBEK protocol provides three security services: Authentication, integrity and no repudiation. The fundamental idea behind providing these services is the watching mechanism. The watching Mechanism requires nodes to store one or more records (i.e. current energy level and Node-Id) to be able to compute the dynamic keys used by the source nodes, to decode packets, and to catch incorrect packets either due to communication Problems or potential attacks. However, there are costs (communication, computation, and storage) associated with providing these services. In reality, applications may have different security requirements. For instance, the security needed by a military application.

C. *Operational mode*

This is one of the operation mode in NEBEK. Here all nodes watch their neighbors, whenever a packet is received from a neighbor node, it is decoded and its authenticity and integrity are verified. Only valid or acceptable packets are forwarded toward the sink. In this mode, a short span of time exists at initial deployment so that no one can hack the network, because it takes time for an attacker to capture a node or get keys. During this period, information to initialize route, may be used by each node to decide which node to watch and a record is stored for each of its one-hop neighbor in its watch-list. To obtain a neighbor's initial energy value, a network-wise master key can be used to transmit this value during this period similar to the shared-key discovery phase of other dynamic key management schemes.

D. Statistical mode

In this operational mode, nodes in the network are configured to only watch some of the nodes in the network. Each node randomly picks node to monitor and stores the corresponding state before deployment. As a packet leaves the source node (originating node or forwarding node) it passes through node(s) that watch it based on probability. Thus, this method is a statistical filtering approach like SEF[7] and DEF[7]. If the current node is not watching the node that generated the packet, the packet is forwarded. If the node that generated the packet is being watched by the current node, the packet is decoded and the plaintext ID is compared with the decoded ID.

Similar to operational mode, if the watcher node wants to forward a packet and it cannot find the key successfully, it will try as many keys as the value of Key Search-threshold before actually classifying the packet as malicious. If the packet is authentic and the current hop is not the final destination then the original packet is forwarded, unless the current node is bridging the network. In the bridging case, the original packet is re encoded with the available bridge energy and forwarded. Since this node is bridging the network, both virtual and perceived energy values are decremented accordingly. If the packet is invalid or unacceptable, which is classified as such after exhausting all the virtual perceived energy values within the virtual Key Search Threshold window, the packet is discarded. This process continues until the packet reaches the sink.

E. Architecture model for NEBEK

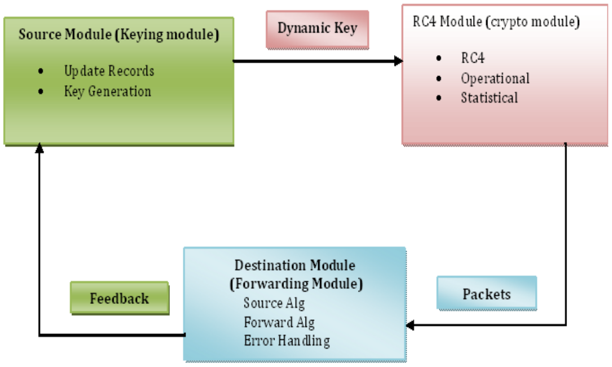


Figure1. Architecture Model for NEBEK

F. Source module (Keying Module)

The Node energy-based source module (keying module) of the NEBEK framework is one of the primary contribution of this project. It is essentially the method used for handling the keying process. It produces a dynamic key that is then fed into the RC4 module (crypto module).

In NEBEK, each node has a certain energy value when it is first deployed in the network. The reasons for using energy as opposed to real battery levels as in the DEEF is that in reality battery levels may fluctuate and the differences in battery levels across nodes may cause synchronization problems, which results in loosing packets.

After deployment, nodes travel across several functional states. The states mainly include node-stay-alive, packet reception, transmission, encoding, and decoding. As each of these actions occurs, the energy in a node is reduced. The current value of the energy, in the node is used as the key to the key generation function. During the initial deployment, each node will have the same energy level, therefore, the initial key, is a function of the initial energy value and an initialization vector. These are pre-distributed to the network. Subsequent keys are the result of the function of current energy and the previous key.

```
Algorithm: Compute Dynamic Key
ComputeDynamicKey(masterkey,packetsiz)
 begin
 j ← temp;
 if  j → 1 then
        K← dynamickey(masterkey,packetsize)
 else
           K← dymamickey( kj-1, masterkey)
 end if
 return K
 end
```

Keying module ensures that each detected packet is associated with a new unique key generated based on the constantly changing value of the energy. After the dynamic key is generated, it is passed to the RC4 encryption module (crypto module), where the desired security services are implemented. The process of key generation is initiated when data is sensed, thus no explicit mechanism is needed to refresh or update keys. Because of the dynamic nature of the keys it makes difficult for attackers to prevent enough packets to break the encoding algorithm.

Each node computes and updates the constantly changing value of its energy after performing some actions. Each action on a node is associated with a certain predetermined cost. Since a node will be either forwarding some other nodes data or injecting its own data into the network, the set of actions and their associated energies for NEBEK includes packet reception, packet transmission, packet encoding, packet decoding energies, and the energy required to keep a node alive in the idle state.

G. RC4 Module (Crypto Module)

The RC4 (Crypto) module uses a simple encoding process, which is essentially the process of permutation of the bits in the packet according to the dynamically created permutation code generated via RC4. The encoding is a simple encryption mechanism adopted for NEBEK. However, NEBEK's flexible architecture allows for stronger encryption mechanisms in lieu of encoding.

In detail:

Due to the resource constraints of networks, traditional digital signatures or encryption mechanisms requiring expensive cryptography is not capable of doing what it is intended to do. The plan must be simple and effective. Thus a simple encoding operation is used [7]. The encoding operation is the process of permutation of the bits in the packet, according to the dynamically created permutation code via the RC4 encryption mechanism. The key to RC4 is created by the previous module (source or keying module). The purpose of the RC4 module is to provide simple confidentiality of the packet header and payload while ensuring the authenticity and integrity of sensed data without incurring transmission overhead of traditional schemes. However, since the key generation and handling process is done in another module, NEBEK's flexible architecture allows for adoption of stronger encryption mechanisms in lieu of encoding.

The packets in NEBEK consists of the ID (i-bits), type (t-bits) (assuming each node has a type identifier), and data (d-bits) fields. Each node sends these to its next hop. The nodes ID, type, and the sensed data are transmitted in a pseudorandom fashion according to the result of RC4.

The RC4 encryption algorithm takes the key and the packet fields (byte-by-byte) as inputs and produces the result as a permutation code shown in the Fig 2. The concatenation of each 8-bit output becomes the resultant permutation code. The key to the RC4 mechanism is taken from the keying module, which is responsible for generating the dynamic key according to the residual energy level.

The resultant permutation code is used in encoding the <ID|type|data> message. Then an additional copy of the ID is also transmitted along with the encoded message. The format of the final packet to be transmitted becomes Packet = [ID,{ID, type, data}k] where {x}k constitutes encoding x with key k. Thus instead of the traditional approach of sending the hash value (e.g., message digests and message authentication codes) along with the information to be sent, we use the result of the permutation code value. When the next node along the path to the sink receives the packet, it generates the local permutation code to decode the packet

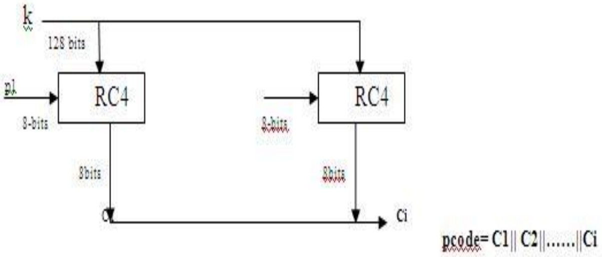


Figure2: RC4 encryption mechanism in NEBEK

Another significant step in the RC4 [8] or crypto module involves how the permutation code dictates the details of the encoding and decoding operations over the fields of the packet when generated by a source node or received by a forwarder node. Specifically the permutation code P can be mapped to a set of actions to be taken on the data stream combination.

The benefits of this simple encoding scheme are:

There is no hash code or message digest to transmit, the packet size does not grow, avoiding bandwidth overhead on an already resource-constrained network, thus increasing the network lifetime.

The technique is simple, thus ideal for devices with limited resources (e.g., PDAs).

The input to the RC4 encryption mechanism, the key, changes dynamically without sending control messages to rekey.

H. The Destination module (Forwarding Module)

The forwarding module handles the process of sending or receiving of encoded packets along the path to the sink.

The final module in the NEBEK communication architecture is the forwarding module. The forwarding module is responsible for the sending of packets (reports) initiated at the current node (source node) or received packets from other nodes (forwarding nodes) along the path to the sink.

The reports traverse the network through forwarding nodes and finally reach the terminating node, the sink. The operations of the forwarding module are explained in this section.

I. Source node Algorithm

When an event is detected by a source node, the next step is for the report to be secured. The source node uses the local virtual energy value and an Initial Vector (or previous key value if not the first transmission) to construct the next key. This dynamic key generation process is primarily handled by the source module. The source module fetches the current value of the virtual energy from the NEBEK module. The key is used as input into the RC4 algorithm inside the RC4 module to create a permutation code for encoding the <ID|type|data> message. The encoded message and the clear text ID of the originating node are transmitted to the next hop (forwarding node or sink) using the following format: [ID, {ID, type, data}Pc], where {x}Pc constitutes encoding x with permutation code Pc. The local virtual energy value is updated and stored for use with the transmission of the next report.

J. Forward node Algorithm

Once the forwarding node receives the packet it will first check its watch-list to determine if the packet came from a node it is watching. If the node is not being watched by the current node, the packet is forwarded without modification or authentication. Although this node performed actions on the packet (received and forwarded the packet), its local virtual perceived energy value is not updated. This is done to maintain synchronization with nodes watching it further up the route.

If the node is being watched by the current node, the forwarding node checks the associated current virtual energy record stored for the sending node and extracts the energy value to derive the key [6]. It then authenticates the message by decoding the message and comparing the plaintext node ID with the encoded node ID. If the packet is authentic, an updated energy value is stored in the record associated with the sending node. If the packet is not authentic it is discarded. The virtual energy value associated with the current sending node is only updated if this node has performed encoding on the packet.

IV. Software Implementation

.Net is used to implement this project. C# is a multi-paradigm programming language encompassing imperative, declarative, functional, generic, object-oriented (class-based), and component-oriented programming disciplines. It was developed by Microsoft within the .NET initiative and later approved as a standard by Ecma (ECMA-334) and ISO (ISO/IEC 23270). C# is one of the programming languages designed for the Common Language Infrastructure.

C# is intended to be a simple, modern, general-purpose, object-oriented programming language. Its development team is led by Anders Hejlsberg. The most recent version is C# 4.0, which was released on April 12, 2010.

Design goals

The ECMA standard lists these designgoals for C#

- C# language is intended to be a simple, modern, general-purpose, object-oriented programming language.
- The language, and implementations thereof, should provide support for software engineering principles such as strong type checking, array bounds checking, detection of attempts to use uninitialized variables, and automatic garbage collection. Software robustness, durability, and programmer productivity are important.
- The language is intended for use in developing software components suitable for deployment in distributed environments.
- Source code portability is very important, as is programmer portability, especially for those programmers already familiar with C and C++.
- Support for internationalization is very important.
- C# is intended to be suitable for writing applications for both hosted and embedded systems, ranging from the very large that use sophisticated operating systems, down to the very small having dedicated functions.

Although C# applications are intended to be economical with regard to memory and processing power requirements, the language was not intended to compete directly on performance and size with C or assembly language.

A. UML diagram

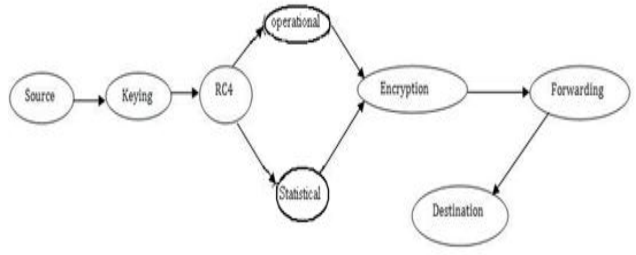


Figure 4: UML Diagram for NEBEK.

Figure 4 shows the UML diagram, various modules used in this project. It shows how the packets have been transferred from source to destination node that is the destination node. The packet will undergo various steps like keying, encoding, etc. Here in this project operational mode and statistical mode will take care of malicious packets.

B. Data flow diagram

The Figure 5 shows the flow of data in NEBEK.

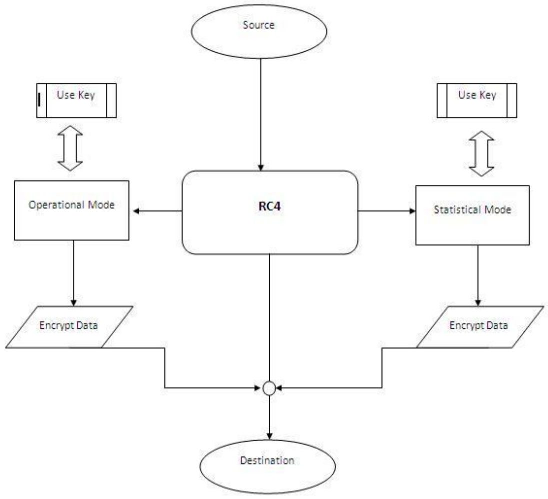


Figure5: Data Flow Diagram

The dynamic key will be generated in the source module and will be sent to RC4 module for encryption along with data to be sent. RC4 module encrypts each packet along with the unique dynamic key for each packet. The encrypted packet will be sent to operational or statistical mode to check the authenticity of the packet. If the packet are valid, they will be forwarded to the destination, where the packet will be decrypted. At the destination again the packet will be checked for authenticity and integrity of the packet.

V. Results And Discussion

Due to the broadcast nature of the networking medium, attackers may try to eavesdrop, intercept, or inject false messages. In this paper, we mainly consider the false injection and eavesdropping of messages from and outside malicious node; hence, insider attacks are outside the scope of this paper. This attacker is thought to have the correct frequency, protocol, and possibly a spoofed valid node ID.

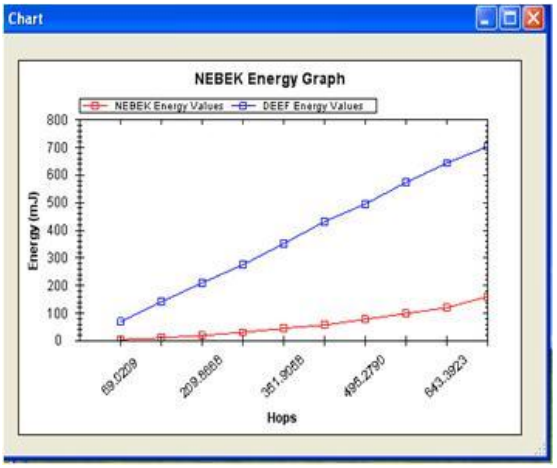


Figure6: Comparison of energy efficiency for NEBEK and DEEF

Filtering efficiency of statistical mode vs. operational mode

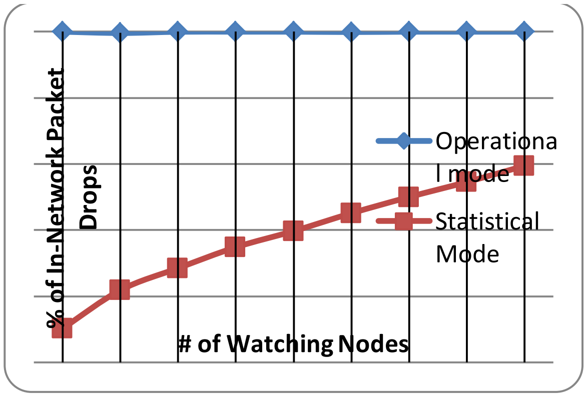


Figure7: Comparison of Modes of NEBEK

In Statistical and operational , in order for an attacker to be able to successfully inject a false packet, an attacker must forge the packet encoding (which is a result of dynamically created permutation code via RC4). Given that the complexity of the packet is 2l, [4]where l is the sum of the ID, TYPE, and DATA fields in the packet, the probability of an attacker correctly forging the packet [6] is:

P forg $= 1/2^{l}$ where l= packetsize

Accordingly, the probability of the hacker incorrectly forging the packet, and therefore, the packet being dropped

Ppdrop = 1- P forg

Since operational mode, authenticates at every hop, forged packets will always be dropped at the first hop with a probability of Ppdrop .

On the other hand, statistical mode, statistically drops packets along the route. Thus, the drop probability for statisticl mode, (Pdrop_II) is a function of the effectiveness of the watching nodes as well as the ability for a hacker to correctly guess the encoded packet structure. Accordingly, the probability of detecting and dropping a false packet at one hop when randomly choosing r records (nodes to watch) is:

P drop_II = $\frac{r}{N} * (1 - Pforg)$

Thus, the probability to detect and drop the packet when choosing r records after h hops is:

P pdrop_II = $1 - (1 - P\ drop_{II})^{h}$

Where h- Number of hops

r- Number of records.

Operational mode is always able to filter malicious packets from the network with its 100 percent filtering efficiency. This is mainly due to the fact that malicious packets are immediately taken out from the network at the next hop. However, the filtering efficiency of Statistical mode is closely related to the number of nodes (r) that each node watches.

VI. Conclusion

Communication is very costly for any network. Independent of the goal of saving energy, it may be very important to minimize the exchange of messages (e.g., military scenarios). To address these concerns, we presented a secure communication framework called Node Energy- Based Encryption and Keying. In comparison with other key management schemes, NEBEK has the following benefits: 1) it does not exchange control messages for key renewals and is therefore able to save more energy and is less chatty, 2) it uses one key per message so successive packets of the stream use different keys—making NEBEK more resilient to certain attacks (e.g., replay attacks, brute-force attacks, and masquerade attacks), and 3) it unbundled key generation from security services, providing a flexible modular architecture that allows for an easy adoption of different key-based encryption or hashing schemes. renewals and is therefore able to save more energy and is less.

References

[1] I.F.Akyildiz, W.Su, Y.Sankarasubramaniam and E. Cayirci, “Wireless Sensor Networks: A Survey”, Computer Networks vol. 38, no. 4, pp. 393-422, Mar. 2002.

[2] C.Vu,R.Beyah and Y. Li, “A Composite Event Detection in Wireless Sensor Networks,” Proc. IEEE Int’l Performance, Computing and Comm. Conf. (IPCCC ’07), Apr. 2007.

[3] G.J. Pottie and W.J. Kaiser, “Wireless Integrated Network Sensors,” Comm. ACM, vol. 43, no. 5, pp. 51-58, 2000 Computerworld.

[4] H. Hou, C. Corbett, Y. Li, and R. Beyah, “Dynamic Energy-Based and Filtering in Sensor Networks”, Proc. IEEE Military Comm. Conf. (MILCOM ’07), Oct. 2007.

[5] Huy Hoang Ngo, Xianping Wu, Phu Dung Le, mpbell Wilson, and Balasubramaniam Srinivasan ,“Dynamic Key Cryptography and Applications,” Monash University,900 Dandenong Road, Caul¯eld East,Victoria, 3145, Australia Feb. 9, 2009.

[6] Raheem A. Beyah, Yingshu Li, John A “Virtual Energy-Based Encryption and Keying for Wireless Sensor Networks”.

[7] Ye, H. Luo, S. Lu, and L. Zhang, “Statistical En-Route Filtering of Injected False Data in Sensor Networks,”.

[8] Allam Mousa and Ahmad Hamad Electrical Engineering Department An-Najah University, “Evaluation of the RC4 Algorithm for Data Encryption “.

AUTHORS PROFILE

Dr.S.Bhargavi is presently working as a Professor in the department of Electronics and Communication engineering, SJCIT, Chikballapur, Karnataka, India. She is having 12 years of teaching experience. Her areas of interest are Robotics, Embedded Systems, Low Power VLSI, Wireless communication, ASIC and Cryptography.

Ranjitha B.T received Bachelor of Engineering degree in Computer Science from Visvesvaraya Technological University, Belgaum, Karnataka, India, in 2008. Currently she is pursuing M. Tech in Digital Communication and Network in

Visvesvaraya Technological University, Belgaum, Karnataka, India. She has 2 Years of Teaching experience. Her areas of interest are Computer Network, Cryptography and Wireless Communication.

Barriers in Adoption of Health Information Technology in Developing Societies

Fozia Anwar
Health Informatics Unit
COMSATS Institute of Information Technology
Islamabad, Pakistan

Azra Shamim
Computer Science Department
COMSATS Institute of Information Technology
Islamabad, Pakistan

Abstract—**This paper develops the conceptual framework of barriers faced by the decision makers and management personnel of health sector. The main theme of this paper is to give a clear understanding about the adaption barriers of health technology faced by developing societies. The information about barriers would be useful for policy makers to decide about the particular technology. So that they can fulfill the defined mission of their organizations. Developing a conceptual framework is the first step in building organizational capacity. Information technology in health sector is spreading globally. Use of health information technology is offering evidence-based practice to endorse health and human prosperity. Globalization of health information system is inevitable for establishment and promotion of healthcare sector in developing societies. Present health systems in developing societies are inadequate to meet the needs of the population. Health sector of developing societies is facing a lot of barriers in establishment and promotion of health information system. These barriers include lack of infrastructure, cost, technical sophistications, lack of skilled human resources and lack of e- readiness of medical professionals. In this paper authors conducted a survey of hospitals in Pakistan to identify and categorized adaption barriers in health information technology. Existing health system should be transformed by using HIT to improve health status of population by eliminating barriers identified in this paper.**

Keywords-component; Health Technolog; Health Information System; Barrier in e- Health.

I. INTRODUCTION

Health information technology (HIT) consists of set of technologies with a great diversion for transmitting and managing health care data for the use of all stakeholders. Stakeholders of health information include payer, providers and all other groups having interest in health and health care system [1].

HIT means the use of computer in the form of physician digital assistance, electronic health records, computerized physician order entry system by doctors, patients, hospitals, laboratories, x-ray facilities and all other stake holders. Health information technology is very important part of constantly changing environment of health care system. With the use of HIT, health related information can be communicated well and can be used in evidence based decision making process [2]. Benefits of HIT include reduced paper work in hospital environment [3] which results in reducing medical errors, minimizing the repetition of medical tests which affects in lowering health care cost [4]. Evidence-based practices play a key role in efficient health information system quality, outcome and efficiency of health. Health care system can be improved by having timely and accurate information for evidence based decisions at all levels [5]. Proactive uses of information technology in health sector empowers consumers of health services as they can easily access health information as well as decision tool and by the employment of HIT, health professionals can collaborate more easily when distance is a major factor [6]. In past HIT applications were used for administrative activities and financial activities rather assisting and delivering health care [7].

The rest of the paper is organized into different sections. Section 2 gives an overview of research methodology. Section 3 presents literature review. Barriers faced by under developed countries in establishment of HIT are discussed in Section 4. Results are presented in section 5 Section 6 discusses suggestions. Section 7 concludes the research work.

II. RESEARCH METHODOLOGY

This paper presents a cross sectional quantitative survey. The focus of authors is to identify potential barriers of information technological interventions in health sector. Sample size is 9 hospitals of Rawalpindi and Islamabad cities; 4 hospitals are from government and 5 are from public sector. Convenience sample technique was used for the study. Sample size will be taken without taking it in consideration that whether they are private or government, teaching or non teaching, profitable or non profitable hospitals. Inclusion criteria is that any hospital having at least two departments computerized or having their departmental information system regardless of their integration within the department or within the hospitals. Data was analyzed by using SPSS. Only percentages and frequencies are used for analysis.

III. LITERATURE REVIEW

A. Health Information Technology (HIT)

Health information system is spreading worldwide and thus promoting health and prosperity for humans. Simplest definition of hospital information system is “Computer application in hospital” [8]. HIS is an N- tier application suit built for a single location or multi location environment [9]. Important features of an effective and functional HIS should include easy, friendly and ready to use, well integrated, customization property and possible tracking and alert facility. Least but not last automation back up is necessary so that no data loss should occur [10].

B. *Health Technology*

The term health technology covers a range of methods used to promote health services, prevent diseases and treat them appropriately. With the use of health technology short-term care (rehabilitation) and long term care which include drugs, devices and procedures can be improved. There are few examples of health information technology used in different health setups which include the computerization of medical records in hospitals and clinics, use of Internet for document delivery, information exchange and communication, development of e-cards for patient identification, development of electronic scheduling system to give appointments, hospital labs and hospital admission examination and computerized protocol for diagnosis and giving treatment support. HIT system provides improved decision-making and appropriate use of diagnostic laboratory tests and therapeutic agents.

IV. Barriers Faced by Underdeveloped Countries in Establishment and Promotion of Hit

Presently health information system infrastructure is deficient in resources to meet the demands and needs of increasing population in developing countries. Health care systems of developing countries have major barriers like poverty and lack of technological sophistication. The basic difficulties or barriers in using information technologies include poor or inadequate infrastructure, insufficient access to the hardware and inadequate or poor resources allocation. By eliminating these barriers population health status can be improved in developing countries. [10]

The authors have identified following barriers in establishment and promotion of HIT in under developed countries. These barriers are divided into six groups i.e. ICT Infrastructure, cost and time, national policies toward HIT, social and cultural, educational, organizational, and ethical barriers.

A. *Barriers Related To Infrastructure*

Most of underdeveloped countries do not have required technological infrastructure to establish national health information system [11] hence cannot promote HIT in private and public hospitals. Reshaping infrastructure of existing health system is very crucial. The following is description of infrastructural barriers:

- Poor or Inadequate Infrastructure

Most developing countries do not have adequate required infrastructure such as computer hardware, software, wired and wireless communication channels, Internet, and skilled professional human resource. The availability and operation of these components of ICT Infrastructure are necessary for establishment and promotion of HIT in under developed countries. Strong infrastructure is required for the strong health information system to improve existing health system by planning and introducing new health care interventions which results in achieving better health goal [12],[13],[14]. There are poor or inadequate resources allocation for implementation and use of the health technology in the developing countries.

- Provision of Computer Hardware and Software

HIT requires specialized software and hardware to improve public health by making evidence based decisions. Often these software and hardware tools are costly and require sufficient training for proper operation.

- Poor Internet Availability

Poor internet availability is a vital infrastructure barrier. Health care specialists have poor access to real time information and the available information is not according to the local situation. This available information cannot be used for evidence based decisions. Without having a proper local area network and internet facility inter-organizational and intra-organizational communication is not possible. This is a backbone for any information system [14], [15].

- Lack of Professional Human Resource Workforce nd Lack of Trainings to Produce this Workforce

A computerized information system requires skilled personnel for its effective operation. Training is one of the aspects for use of any new technology. Deficiency of skilled workforce can be overcome by providing appropriate training in the required area. A proper training module in constructing architecture of a reliable database should be available. If it is not implemented then outcomes or results gained by such type of databases gives unauthentic results which can neither be used for decision making process nor for evidence based practice. Training requires cost as well as time.

B. *Cost and Time Barriers*

Major problem in organizing workshops and trainings for establishment and implementation of HIT in under developed countries is financial and time constraint [14], [16], [17], [18]. Transformation of any system is a difficult task and cannot complete in short time period. Barriers like lack of skilled workforce, infrastructure, and cost along with other effects like initial decrease of productivity due to adjustment with new technological environment and system itself impose, strong limits to the introduction and adaption of new health technologies [14], [18]. It requires years and years for transformation process to complete.

C. *National Policies towards HIT*

Efficient, effective and secure national policy can address the local health needs according to the changing environment is needed. These policies can be devised by policy makers and practitioners to assess and implement research evidences [14], [16]. Enforcing the legislation is difficult in developing countries and acceptance by the community for the transformation of any system is harder [14], [20]. Total amount of Rs. 663 billion has been allocated in PSDP (Public Sector Development Program) 2010-11 for various ongoing and new schemes Out of this only Rs.16944.5 million gas been allocated for development of Health Division with 2.1% population average annual growth which is abysmally low [21].

D. Social and Cultural Barriers

Digital divide and e-readiness are major social and cultural barriers in establishment and use of health information system. These barriers include lack of stakeholder's interest, less motivation, anxiety to adapt and use new technology. Health care personnel are difficult to convince for use of new health technologies. As they are more comfortable with their conventional approach and routine practice so it is complicated to transform health information system from paper based to digital format [14], [22].

E. Educational Barriers

Professional education in health informatics is badly ignored and missing in curriculum of medical institutes for undergraduates. Although module of education related to IT use in research is included in postgraduate curriculum but it is the need of the hour to include this area in medical professional education at graduate level. Transformation of our existing paper based health system into computerized information system is not possible without providing the basic IT knowledge to health professionals.

F. Organizational Barriers

Organizations and people play a very critical role in implementing and transformation of an information system. First of all there are no documented studies available regarding level and use, benefits, cost, risk analysis and other aspects of health technology in health sector of underdeveloped countries and if they are available for the developing countries they are not well communicated. Secondly, people at higher positions and posts, whose needs of reporting are adequately being catered by the existing system, do not favor HIT as they think that the employment of new technology is wastage of both the money and time. Hospitals must address the apprehension of physicians because if by using HIT their professional responsibilities become difficult they will never support its use [23].

V. Results and Discussions

Barriers are classified into different 7 categories. These 7 categories were further explored in context with:

- Important issues decisive in adapting new technology
- Factors not recommending new technology
- Factors affecting adaption of technology despite recommendation of the management
- Factors influencing making decisions about adaption of technology

In figure 1 initial cost barrier is shown. This is a barrier for both private and government sector but private hospitals gave more emphasis on initial cost barrier.

Figure 2 despite ongoing costs analysis and concludes that both private and government hospitals consider it as a barrier but again private hospitals give more importance to this barrier as they are strongly agree that cost is an issue.

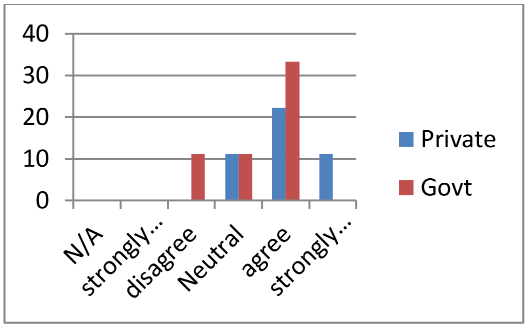


Figure 1. Initial Cost Barrier

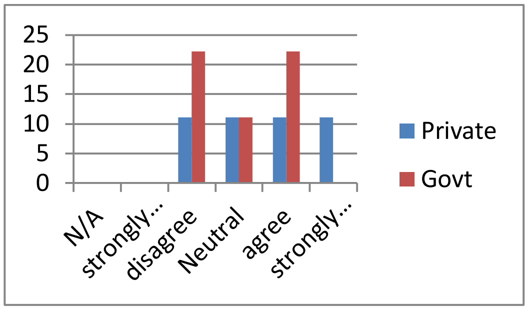


Figure 2. Maintenance Cost

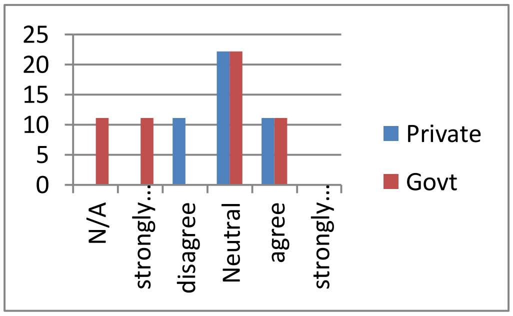


Figure 3. Unproven Returns on Investments

In figure 3 profiles of unproven returns on investment is shown. Both public and private sector agree at the same level. Documented proofs about the success and investment return are not available and even if they are available they are not well communicated to the decision making authorities and stakeholders. Total cost for the establishment of health information system includes acquisition, infrastructure, implementation, maintenance and training cost as shown in the following equation:

$$COST_T=COST_A+COST_{IF}+COST_{IM}+COST_M+COST_T$$

Where

$COST_T$ is the total cost

$COST_A$ is acquisition cost

$COST_{IF}$ is infrastructural cost

$COST_{IM}$ is implementation cost

$COST_M$ is maintenance cost

$COST_T$ is training cost

Furthermore acquisition cost includes cost of hardware and software as shown in the following equation:

$COST_{A=}COST_{H}+COST_{S}$
$COST_{H}$ is hardware
$COST_{S}$ is software

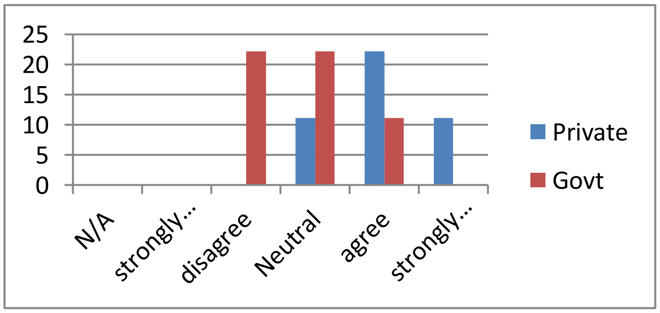


Figure 4. Acceptance by Clinical Staff

In figure 4 behavioral barriers in the form of acceptance by the clinical staff is shown. Private sector seems to be more sensitive to this barrier at high level as compared to the government sector. The reason might be that in government sector all clinical and administrative staff must have to follow the commands as they have no other options while in private sector they can switch to another setup easily.

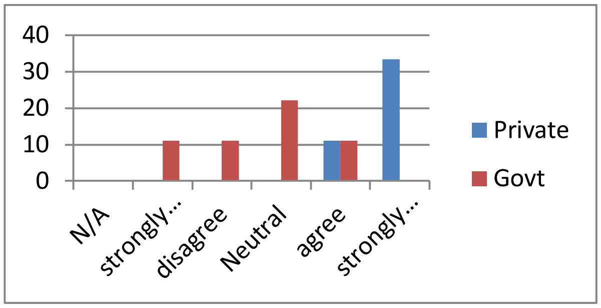


Figure 5. Unavailability of Well Trained IT Staff

Figure 5 shows the unavailability of well-trained IT staff as a barrier in adaption of health technology.

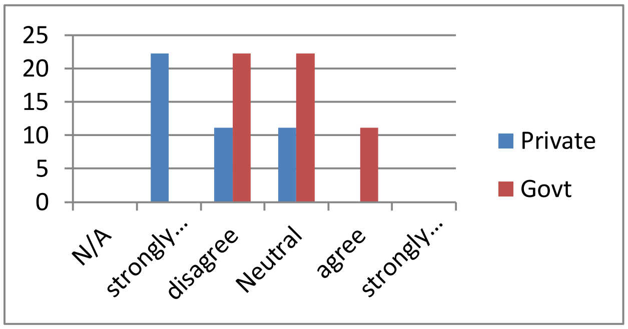


Figure 6. Difficult to Use

In figure 6 barriers regarding the difficulty to use new technology is shown. Only government sector agrees that this is a barrier for adaption of new health technology. Reason behind is that government employees have security about their jobs while personnel in private sector have to update their skills and knowledge accordingly. Interoperability of existing and new system is shown in figure 7.

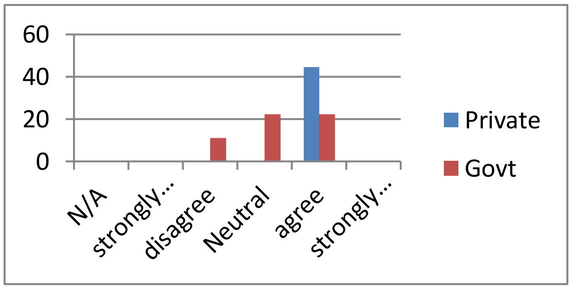


Figure 7. Interoperability with Current System

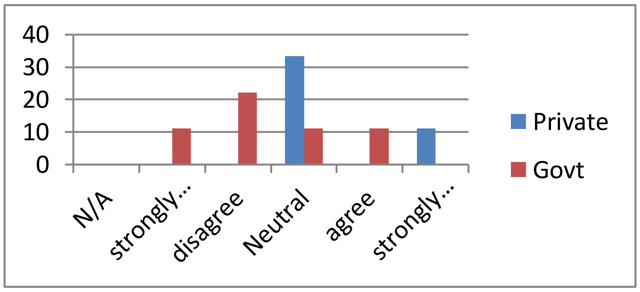


Figure 8. Not Enough Time for Training

As hospitals have a great workload of outpatient and inpatient so sparing time for training to adapt new health technology is a major concern as shown in figure 8.

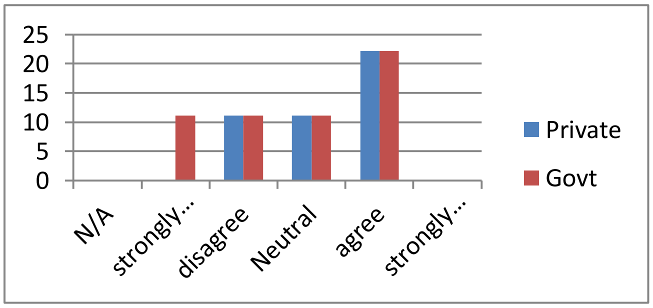


Figure 9. Difficulty in Changing Work Flow Pattern

Figure 9 shows change in current workflow pattern as a barrier. This is a very important aspect for both public and private sector in adaption of any new health technology. Any small change in workflow due to adaption of new health technology affects more on the clinical and administrative functions of the hospital. Change in current workflow pattern of any hospital cannot be made in small time period.

In figure 10 helping factors in making decision about adaption of technology are categorized. Emphasis by ministries or government policies, knowledge translation and feedback are the important factors which help in implementing new technologies in hospitals. Next important helpful factor is stage or phase adaption then factors physician demand and transparent funding formula come. Physician will be demand new health technology in their working environment only when knowledge translation will be improved and they will aware of the benefits of new health technology. And emphasis by ministry or authorities is an important helpful factor considered.

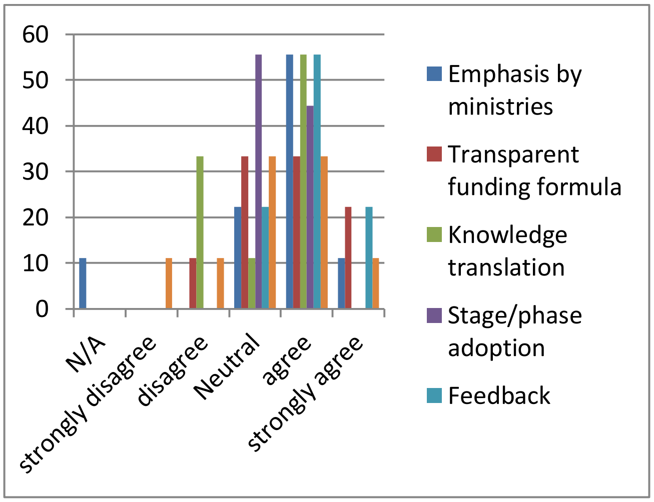


Figure 10. Factors Influencing Making Decision about Adaption of Technology

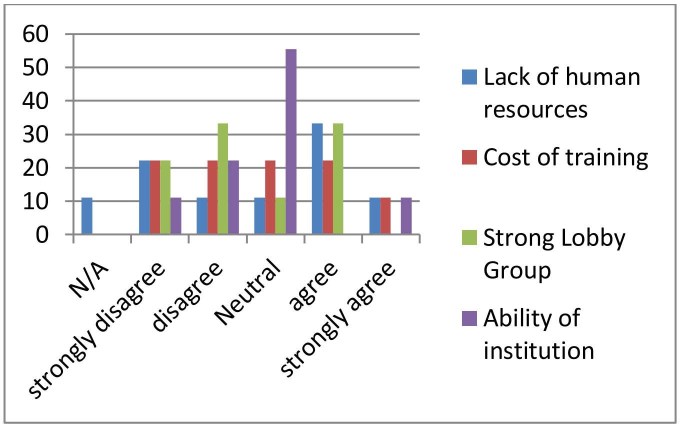


Figure 11. Factors Affecting Adaption of Technology despite Recommendation of the Managment

In figure11 barriers are categorized according to their effect in adaption of technology despite recommendation of the management. Strong lobby group of the health personnel due to many authentic reasons is one of the main factors. The other factor is lack of qualified and trained human resource to operate and maintain new health technology. Cost of training is also a barrier for management to adapt new health technology.

VI. Suggestions

Fixed and implicit percentage of the total budget of hospital should be allocated for the health informatics or health / IT department. Good collaboration is strongly required between health informatics personnel and professional group for a better understanding. So a group of people should be trained in both the fields of informatics and health. Thus they can act as a bridge between these two professions.

The government, ministry of health or other stakeholders might impose the important measures for adaption of health technology to promote HIT. New flexible methods for implementations should be adapted. Phase adaption is recommended. Initially the existing system and new technology should be run side by side. Hundred percent changes in the system track sometimes lead to failure.

VII. Conclusion

Adaption barriers of new health technology are hot issue in developing societies. Developed countries have adapted new health technologies in their health sector to a great extent. Developing countries are still struggling for it and facing a lot of barriers like lack of infrastructure, cost, time and benefit analysis, lack of skilled workforce, national policies and motivation of health related personnel. It is high time to recognize that evidence-based informatics is very helpful in putting science into the field or practice. Enough professional skills related to HIT are lacking. HIS cell or health technology projects are lacking in budget. The time period, which is required for well implementation of the specific health technology, is quite long.

As Informatics is a constantly changing field and technology is advancing day by day, so when you implement one technology the new version or latest technology comes. System up-gradation along with this change is unavoidable. Complexity in maintaining, integrating and interoperability with the existing system make the use of health technology very hard and difficult. These barriers are main hurdle in adaption of health technologies and by addressing these barriers adaption process can be successful.

References

[1] David Blumenthal and John Glaser, Information Technology Comes to Medicine" , New England Journal of Medicine Vol. 356, 24th , June 2007

[2] Innovators and Visionaries: Strategies for Creating a Person-centered Health System. FACCT: Foundation for Accountability; September 2003.

[3] Health Information Technology, Can HIT Lower Costs and Improve Quality? 2005 RAND Corporation. RAND Health research reported http://www.rand.org/pubs/research_briefs/RB9136/index1.html

[4] Using Health Information Technology to Manage Your Information Computers Can Help You and Your Doctor Manage Your Health Care, Michael Bihari, MD, former About.com Guide Updated April 06, 2010

[5] http://www.thefreedictionary.com/health+information+system, Accessed Date: June 2010

[6] Innovation of health technology, Report by centre of AHIP centre of policy and research, Effective new solution for Americans health insurance plans

[7] Audet AM, Doty MM, Peugh J, Shamasdin J, Zapert K, Schoenbaum S, Information technologies: when will they make it into physicians' black bags? MedGenMed. 2004; 6:2. [PMID: 15775829](13)

[8] health information support, regional office for the eastern Mediterranean, WHO, Cairo Egypt, May, 2001

[9] www.hospitalinformationsystem.com visited on 16th April, 2011

[10] Laurence J. Krieg, Introduction to hospital information system module M30c, Washtenair community college, 26 march1999.

[11] www.who.int/topics/millennium...goals/.../ICT_for_Accountability.pdf

[12] Azubuike M. C. J.E. Ehiri, Health information systems in developing countries: benefits, problems, and prospects, The Journal of the Royal Society for the Promotion of Health, Vol. 119, No. 3, pp 180-184 (1999), DOI: 10.1177/146642409911900309, © 1999 Royal Society for the Promotion of Health,

[13] Lansang MA, Dennis R, Building capacity in health research in the developing world, Bull World Health Organ. 2004 Oct;82(10):764-70, Review, PMID: 15643798 [PubMed – indexed for MEDLINE]

[14] Mugdha R Oak, A review on barriers to implementing health informatics in developing countries, Journal of Health Informatics in developing countries, vol 1, No. 1, 2007

[15] Sluijs MB, Veeken H, Overbeke AJ., Deficient information in developing countries: Internet alone is no solution, Ned Tijdschr Geneeskd. 2006 Jun 17; 150(24):1351-4, Dutch, PMID: 16808368,

[16] Chinnock Paul, Global review on access to health information in developing countries, Discussion Paper 1. The role of systematic reviews, Cochrane Collaboration, http://www.hi-urope.co.uk/files/2004/9962.htm

[17] Garner Paul et. al., Getting research findings into practice Implementing research findings in developing countries, BMJ 1998, pp 531-535, 22nd August, 1998 ,

[18] McDonald Michael D., Health Information Infrastructure in Developing Countries, Global Health Initiatives, Inc., http://www.greenstar.org/GHI/Developing%20Countries.htm

[19] Martínez A, et al. Analysis of information and communication needs in rural primary health care in developing countries, IEEE Trans Inf Technol Biomed. 2005

[20] Tan Tessa, Torres Edejer. Information in practice Disseminating health information in developing countries: the role of the Internet, BMJ 2000; 321, pp 797-800, 30th September 2000

[21] Pakistan budget 2010-2011 http://www.onepakistan.com/finance/budget/index.php Accessed on 12th April 2011.

[22] Tan Tessa, Torres Edejer. Information in practice Disseminating health information in developing countries: the role of the Internet, BMJ 2000; 321:797-800, 30 September 2000

[23] Reider J. Cedars-Sinai Medical Center suspends CPOE, January 2003. (Accessed Date May 24, 2007, Available at http://www.docnotes.net/000866.html.)

Implementation of Locally Weighted Projection Regression Network for Concurrency Control In Computer Aided Design

A.Muthukumaravel
Research Scholar,
Department of MCA,
Chennai – 600 117, India.

Dr.S.Purushothaman
Principal,
Sun College of Engineering
and Technology,
KK District – 629 902, India

Dr.A.Jothi
Dean,
School of Computing Sciences,
VELS University,
Chennai – 600 117, India.

Abstract—**This paper presents implementation of locally weighted projection regression (LWPR) network method for concurrency control while developing dial of a fork using Autodesk inventor 2008. The LWPR learns the objects and the type of transactions to be done based on which node in the output layer of the network exceeds a threshold value. Learning stops once all the objects are exposed to LWPR. During testing performance, metrics are analyzed. We have attempted to use LWPR for storing lock information when multi users are working on computer Aided Design (CAD). The memory requirements of the proposed method are minimal in processing locks during transaction.**

Keywords-*Concurrency Control; locally weighted projection regression*; T*ransaction Locks; Time Stamping.*

I. INTRODUCTION

Concurrency control is one of the essential characteristics of transaction management to ensure consistency of database. Maintaining consistency in transactions of objects is mandatory. During computer aided design (CAD), many people will be accessing different parts of same objects according to the type work allotted to them. As all the parts of the same objects are stored in a single file, at any point of time, there should not be corruption of data, inconsistency in storage and total loss of data.

The concurrency control requires proper locking methods for controlled transactions. The most common way in which access to objects is controlled by 'locks'. In a database operation, lock manager plays an important role. It checks whether one or more transactions are reading or writing any object 'I' where 'I' is an item. It is the object of that record, for each item I. Gaining access to I is controlled by manager and ensure that there is no , access (read or write) would cause a conflict. The lock manager can store the current locks in a lock table which consists of records with fields (<object>, <lock type>, <transaction>) the meaning of record ('I', 'L', 'T') is that transaction 'T' has a lock of type 'L' on object 'I '[1-4].

The process of managing simultaneous operations on the database without having them interfere with one another is called concurrency. [5-8] When two or more users are accessing database simultaneously, concurrency prevents interference. Interleaving of operations may produce an incorrect result even though two transactions may be correct. Some of the problems that result in concurrency [11-20] are lost update, inconsistent analysis and uncommitted dependency.

II. PROBLEM DEFINITION

There is inability to provide consistency in the database when long transactions are involved. It will not be able to identify if there is any violation of database consistency during the time of commitment. It is not possible to know, if the transaction is with undefined time limit. There is no serializability when many users' work on shared objects. During long transactions, optimistic transactions and two phase locking will result in deadlock. Two phase locking forces to lock resources for long time even after they have finished using them. Other transactions that need to access the same resources are blocked. The problem in optimistic mechanism with 'Time Stamping' is that it causes repeated rollback of transactions when the rate of conflicts increases significantly. Artificial neural network [9] with locally weighted projection regression (LWPR) has been used to manage the locks allotted to objects and locks are claimed appropriately to be allotted for other objects during subsequent transactions.

Inbuilt library drawing for the dial of fork (Figure 1) is available in AutoCAD product. The fork is used in the front structure of a two wheeler. Due to customer requirements, the designer edits the dial of fork in the central database by modifying different features. Consistency of the data has to be maintained during the process of modifications of different features. A specific sequence of locking objects has to be done whenever a particular user accesses a specific feature of the dial of fork. Each feature is treated as an object. The features are identified with numbers. In this work, O1 refers to an object / feature marked as 1. The major objects involved in creating the dial of the fork are hollow cylinder, wedge and swiveling plate. The various constraints that have to be imposed during modifications of features by many users on this dial of fork are as follows:

- During development of features, hollow cylinder details should not be changed.

- External rings are associated with hollow cylinder.
- The circular wedge has specific slope and associated with hollow cylinder.

This dial of fork has following entities.

1) Features 1, 2, (set 1)
2) Features 10, 11,12,13,14 (set 2)
3) Features 5,6,7,8 (set 3)
4) Features 3, 4 (set 4)

Set 1, set 2, set 3, set 4 can be made into individual drawing part files (part file 1, part file 2, part file 3 and part file 4) and combined into one assembly file (containing the part files 1,2, 3 and 4 which will be intact). When the users are accessing individual part files, then transactions in part file 1 need not worry about the type of transactions in part files 2,3,4 and vice versa among them. When the part files 1, 2, 3 and 4 are combined into a single assembly file, then inconsistency in the shape and dimension of the set 1, set 2, set 3 and set 4, during matching should not occur. Provisions can be made in controlling the dimensions and shapes with upper and lower limits confirming to standards. At any time when a subsequent user is trying to access locked features, he can modify the features on his system and store as an additional modified copy of the features with Time Stamping and version names (allotted by the user / allotted by the system).

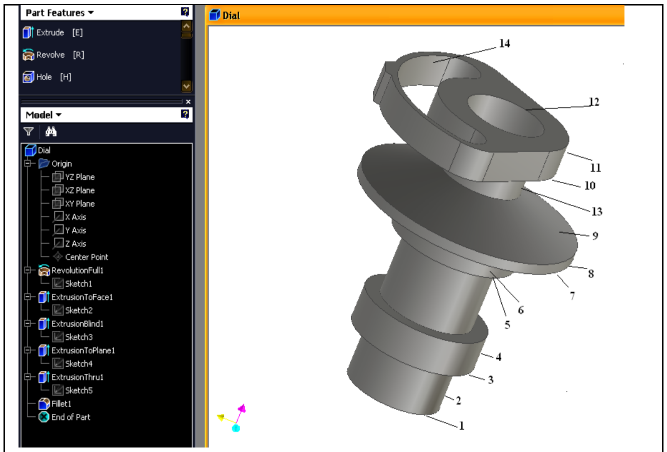


Figure 1. Dial of fork

1 Lower end, 2 Height of the end part, 3 External support, 4 Height of the external support, 5 Support for the wedge, 6.Height of the support for the wedge, 7 Wedge, 8.Thickness of the wedge, 9 Slope of the wedge, 10.Wedge lock, 11.Height of the wedge lock, 12 Concentric hole, 13. Separator, 14.Guideway

III. LOCALLY WEIGHTED PROJECTION REGRESSION

Locally Weighted Projection Regression (LWPR) [10] is an algorithm that achieves nonlinear function approximation in high dimensional spaces even in the presence of redundant and irrelevant input dimensions. At its core, it uses locally linear models, spanned by a small number of univariate regressions in selected directions in input space. The nonparametric local learning system

i. learns information rapidly with second order learning methods based on incremental training.

ii. uses statistically sound stochastic cross validation to learn information.

iii. adjusts its weighting kernels based on local information only.

iv. has a computational complexity that is linear in the number of inputs, and

v. can deal with a large number of possibly redundant and irrelevant – inputs.

The structure of the event loop is shown in Fig.2. The algorithm is at one of the four action states at any given point of time. The INITIALIZE phase is used to initialize the LWPR and read in the script variables from the script file and fill in default values for those variables not specified in the script file.

The TRAIN phase of the algorithm draws data from the training data set file and trains the local model on it. After every 'evaluation', the program goes into the EVALUATE phase where the learned model is tested against the novel (test) data set.

When the number of iterations has exceeded the 'max_iterations' count or the change of normalized mean squared error (nMSE) between the last EVALUATE phase and the current, a checking is done to find if the values is below a THRESHOLD. In such case, the program goes into the RESULT phase during which, it saves the learned LWPR.

A. *Adding an extra projection dimension*

The program can be initialized with 'init_n_reg=1'. It is taken as a special case of one projection LWPR where the number of local projection employed by each local model is restricted to one. The distance metric adjusts in order to accommodate for this restrictions. This may take a long time to converge for regression problems in which the inherent dimensionality is high. If there is no prior knowledge of the intrinsic local dimensionality, then initialize the variable 'init_n_reg=2'. This implies one dimension for the regression and the additional dimension used to check whether to add an extra projection or not. The mean squared error of the current regression dimension is compared against the previous one. Only if the ratio $\frac{\mathrm{nMSEr}-1}{\mathrm{nMSEr}} < \mathrm{add\_threshold}$, a new regression is added. This criterion is used in conjunction with a few more checks to ensure that the decision is based on enough data support before deciding to add a new dimension.

B. *Pruning an existing RF (local model)*

If there is a local model that elicits substantial activation in response to a training data, it prevents the allocation of an additional local model for that training data. Since the distance metric is changing with the gradient descent updates, there can arise cases in which there is a considerable overlap between two local models. Due to this pruning of the variables required. The pruning is due to

i. too much overlap with another one (If 2 receptive fields elicits response greater than 'w_prune' to a training data), then one with the larger error is pruned.

ii. too high variance in the error compared to the 'std' of all the random field(RFs) (determined by the variable 'factor_prune').

iii. excessive error in one of the RFs.

C. Maintaining the local nearest neighbor (nn) list

When using the regression analysis in applications where the input values change smoothly, it is useful to keep a neighborhood list and perform training by looking at only the neighboring local models which are close to each other or have a substantial overlap in their activation profiles. This saves a lot of computing resources as opposed to going through all the local models and finding out those that have enough activation to be updated. It suffices to look at the neighborhood list to check for activations that are above the threshold and need to be updated.

D. Second order updates

The gradient descent updates of the distance metric is speeded up for faster convergence - and is more efficient if Newton's second order gradient information (meta learning) is used. If the 'allow_meta_learning' variable is TRUE, then the second order learning is on.

E. Forgetting factor

The forgetting factor is a variable that is used to discount the effects of the statistics computed at an earlier stage and give more weight to the recent statistics - which are a result of having experienced more data points. It can be thought of as a sliding window over which the stochastic sufficient statistics are accumulated. The forgetting factor ('lambda') takes a value [0, 1] where 0 corresponds to using only the current point and 1 corresponding to not `forgetting 'anything. An annealed forgetting rate can be used which forgets more at the start specified by 'init_lambda' and anneals towards a value closer to one ('final_ lambda') - not forgetting anything based on annealing factor 'tau_lambda'.

IV. Results And Discussion

Let us assume that there are two users editing the dial of the fork. User1 edits O1 and hence O2 will be locked sequentially (Table 1). Immediately user2 wants to edit O2, however he will not get transaction as already O2 is locked. However, user2 or any other user can try to access O3 to O14. The variables used for training the ANN about locks assigned to different objects are transaction id, object id, lock mode (Table 2).

TABLE I. SHAPE AND DIMENSION CONSISTENCY MANAGEMENT

Group	*First feature*	*Remaining feature to be locked*
G1	1	2
G2	10	11,12,13,14
G3	5	6,7,8
G4	3	4

Transaction id represents the client or any other intermediate transactions. Object id represents the entire feature or an entity in the file. Mode represents type of lock assigned to an object.

In Table 2, column 1 represents the lock type. column 2 represents the value to be used in the input layer of the LWPR. Column 3 gives binary representation of Lock type to be used in the output layer of LWPR. The values are used as target outputs in the module during lock release on a data item.

TABLE II. BINARY REPRESENTATION OF LOCK TYPE

Lock type	*(Input layer representation numerical value).*	*Binary representation in target layer of the LWPR*
Object Not locked	0	000
S	1	001
X	2	010
IS	3	011
IX	4	100

Initially, user 1 and user 2 have opened the same dial of fork file from the common database. The following steps shows sequence of execution and results

T1 edits O1 with write mode. Table 4 shows pattern formed for the training.

TABLE III. FIRST TIME PATTERN USED FOR TRAINING LWPR

Object number	*Input pattern*	*Target output pattern*
O_1	[1 1]	[0 1 0]

Step 1: The transaction manager locks objects mentioned in the third column of Table 1. Repeat step 1 with the patterns given in Table 4.

TABLE IV. ADDITIONAL PATTERNS USED FOR TRAINING OML LWPR

Object number	*Input pattern*	*Target output pattern*
O_1	[1 1]	[0 1 0]
O_2	[2 1]	[0 1 0]

Step 2: A new transaction T2 access O2. A pattern is formed to verify if lock has been assigned to O2 and its associated objects O1. Only when the locks are not assigned to O2 and O1 then T2 is allowed.

The following input patterns are presented to the testing module to find if the output [0 0 0] is obtained in the output layer. During testing, the final weights obtained during training will be used. Otherwise it means that lock has been assigned to either O2. In such case, transaction is denied for T2. Else the following Table 5 is presented in step 1.

TABLE V. ADDITIONAL PATTERNS USED FOR TRAINING LWPR

Object number	***Input pattern***	***Target output pattern***
O_1	[1 1]	[0 1 0]
O_2	[2 1]	[0 1 0]
O_3	[3 1]	[0 1 0]
O_4	[4 1]	[0 1 0]
O_5	[5 1]	[0 1 0]
O_6	[6 1]	[0 1 0]
O_7	[7 1]	[0 1 0]
O_8	[8 1]	[0 1 0]

Step 3: To know the type of lock value assigned to an object and for a transaction, testing is used. Testing uses the final weights created by training. The proposed LWPR for lock state learning and lock state finding have been implemented using Matlab 7.

The performance of the LWPR algorithm has been presented on the following criteria.

1. Locking time for each object. (Figure 3)
2. Releasing time for each object.(Figure 4)
3. Total Locking time for each transaction group.(Figure 5)
4. Arrival rate (Figure 6)
5. Response time (Figure 7)

V. CONCLUSION

An artificial neural network with LWPR has been implemented for providing concurrency control to maintain consistency in the CAD database. A dial of fork has been considered that contains 14 objects. The 14 objects have categorized into 4 groups. The transaction behavior and concurrency control by the two users on the 14 objects have been controlled using LWPR network. The LWPR method requires less memory based on the topology used for storing objects and its transactions when compared with conventional method.

REFERENCES

[1] Rosenkrantz .D , Stearns R .and. Lewis P, “System-level concurrency control for distributed database systems,” In ACM Transactions on Database Systems, vol. 3, No. 2, . 1978, pp. 178-198,

[2] Peter A. Buhr, Ashif S. Harji, Philipp E. Lim and Jiongxiong Chen,‘Object-oriented real-time concurrency’, In Proceedings of the 15th ACM SIGPLAN conference on Object-oriented programming systems, languages and applications, 2000,pp. 29-46.

[3] Mihalis Yannakakis, ‘Issues of correctness in database concurrency control by locking’, In Proceedings of the thirteenth annual ACM symposium on Theory of computing, 1981,pp. 363–367.

[4] Klahold .P , Schlageter G. and Wilkes W., August, ‘A General Model for Version Management in Databases’, In Proceedings of the International Conference on Very Large Data Bases, Kyoto, 1986,pp. 319-327.

[5] Katz R.H. and Lehman T.J, ‘Database Support for Versions and Alternatives of Large Design Files’, In IEEE Transactions on Software Engineering, vol. 10, No. 2, ., March 1984,pp. 191-200.

[6] Herrmann U., Dadam P., Küspert K., Roman E. A. and Schlageter G., 1990, ‘A lock technique for disjoint and non-disjoint complex objects’, In Springer Advances in Database Technology — EDBT '90, vol. 416, pp. 219-237.

[7] Garza J. and Kim W., ‘Transaction management in an object-oriented database system’. In Proceedings of the ACM SIGMOD International Conference on the Management, Vol. 17,No. 3,1988,,pp. 37-45.

[8] Eliot B. Moss, , ‘Transaction Management for Object-Oriented Systems’,In Proceedings of the IEEE Computer Society International Workshop on Object-Oriented Database Systems, 1986,pp. 229.

[9] Raviram P., Wahidabanu R. S. D. and Purushothaman S., **“**Concurrency Control in CAD with KBMS using Counter Propagation Neural Network**”,** IEEE International Advance Computing Conference, 6-7 March 2009, pp. 1521-1525.

[10] Purushothaman S., Elango M.K. and Nirmal Kumar S., Application of Hilbert Huang Transform with Locally Weighted Projection Regression Method for Power Quality Problems, International Review on Electrical Engineering, vol. 5. no. 5, October 2010, pp. 2405-2412

[11] Mohammed Khaja Nizamuddin, Dr. Syed Abdul Sattar, 2010, “Data Count Driven Concurrency Control Scheme with Performance Gain in Mobile Environments” in Journal of Emerging Trends in Computing and Information Sciences, vol 2 ,no. 2, pp 106-112.

[12] Salman Abdul Moiz, Dr. Lakshmi Rajamani, “An Algorithmic approach for achieving Concurrency in Mobile Environment”, INDIACom, 2007,pp.209-211,.

[13] K M Prakash Lingam, 2010, “Analysis of Real-Time Multi version Concurrency Control Algorithms using Serialisability Graphs” International Journal of Computer Applications (0975 - 8887), vol 1 ,no. 21, pp. 57 – 62.

[14] Quilong Han,Haiwei pan, “A Concurrency Control Algorithm Access to Temporal Data in Real-Time Database Systems,” imsccs, , International Multi symposiums on Computer and Computational Sciences,2008, pp.168-171.

[15] Tae-Young Choe, 2008, “Optimistic Concurrency Control based on Cache Coherency in Distributed Database Systems” IJCSNS International Journal of Computer Science and Network Security, vol.8, no.11, November 2008, pp 148–154

[16] Mohammed Khaja Nizamuddin, Syed Abdul Sattar, “Adaptive Valid Period Based Concurrency Control Without Locking in Mobile Environments” In: Recent Trends in Networks and Communications, Springer CCIS, vol.90, Part 2, , 2010, Springer Heidelberg, pp 349-358

[17] Arun Kumar Yadav & Ajay Agarwal, “An Approach for Concurrency Control in Distributed Database System” in International Journal of Computer Science &Communication, vol. 1, no. 1, January-June 2010, pp. 137-141

[18] Arumugam.G and Thangara.M "An Efficient Locking Model For Concurrency Control In Oodbs" Data Science Journal, vol 4, 31 August 2005, pp 59-66.

[19] Poonam Singh, Parul Yadav, Amal Shukla and Sanchit Lohia "An Extended Three Phase Commit Protocol for Concurrency Control in Distributed Systems" in International Journal of Computer Applications (0975 – 8887) vol 21,no.10, May 2011

[20] Jinhua Guo, "An Exploratory Environment for Concurrency Control Algorithms" in International Journal of Computer Science vol 1 ,no 3, March 29, 2006, pp 203-211.

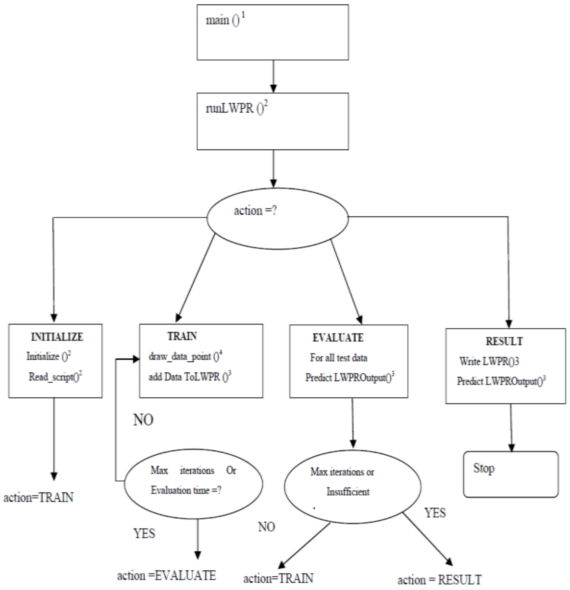


Figure 2. LWPR training modules

Filenames: 1.lwpr_main(), 2.lwpr_test(), 3.lwpr(), 4.utilities()

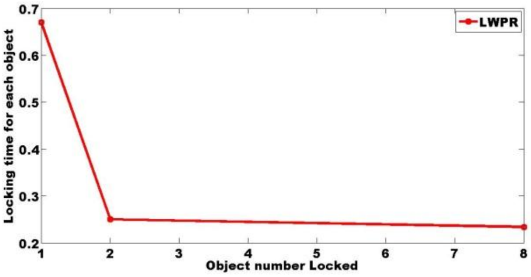


Figure 3. Locking time for each object

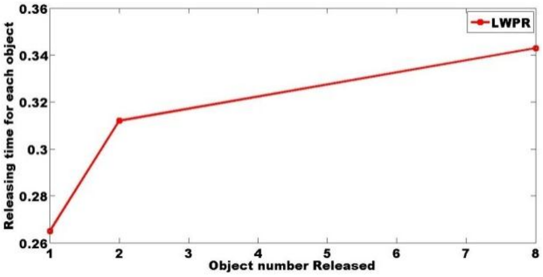


Figure 4. Releasing time for each object

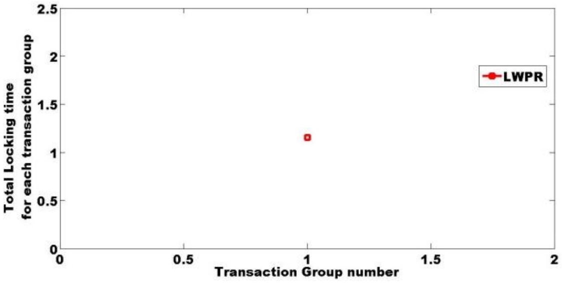


Figure 5. Locking time for each group

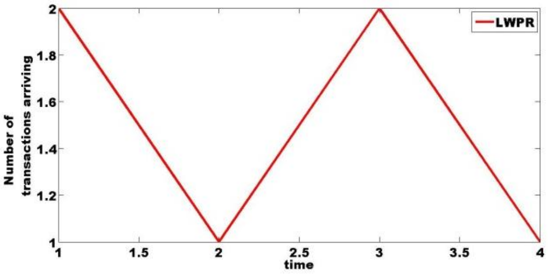


Figure 6. Arrivals rate

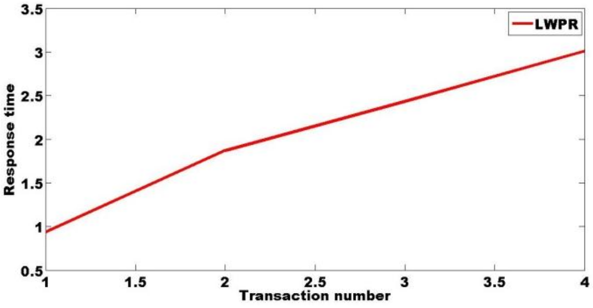


Figure 7. Response time

Transforming Higher educational institution administration through ICT

J. Meenakumari
Associate Professor, Christ University
Bangalore, India

Dr. R. Krishnaveni
Professor, PSG Institute of Management
Coimbatore, India

Abstract—**The rapid development in Indian higher education sector has increased the focus on reforms in higher educational institution administration. Efficiency and accountability have become important elements, and the integration of Information and Communication Technology (ICT) into the educational administration process has become a necessity. The objective of this study is to know the current extent of ICT integration in Indian higher education institutions. The factors contributing to the successful integration of ICT into higher education administration (i.e., Knowledge administration and Information administration which constitute e-administration) are also discussed**

Keywords*-Higher education; Information and Communication Technology; Integration; Knowledge administration; Information administration; e-administration.*

I. Introduction

A. Overview

Developing nations has seen a tremendous growth in higher education sector. India is a developing nation and it has been expanding in every field. Education saw an essential change gradually after Independence. Development of education in India brought about a transformation and the concept of education got modified. India has the third largest higher education system in the world, behind only the United States and China (Harsh V. Pant 2008). As on 31.12.2009, there are 504 Universities and University-level institutions including 40 Central Universities, 243 State Universities, 53 State Private Universities, 130 deemed Universities and 33 Institutes of national importance (Ministry of HRD, India).

B. Indian Higher Education System

The department of higher education is an imperative segment of the Ministry of Human Resource Development, India. The ministry is headed by the Minister of HRD. He is currently assisted by two Ministers of State. The Minister provides policy and overall leadership to the Ministry. At the executive level, the department of higher education is headed by a secretary, who is assisted by one additional secretary, and several joint secretaries or equivalent officers. Each joint secretary heads a bureau and at present, work of the department is divided into six bureaus. The higher education annual report of the year 2009-2010 revealed that 136.42 lakh students were enrolled in the institutions of Higher Education as against 123.77 lakhs in the previous year. This clearly shows that the higher education sector is expanding in all areas.

According to the World Bank Report (1994), higher education is of paramount importance for social and economic development of a nation. This rapid development and expansion in higher education sector has necessitated a good Information and Communication technology (ICT)-based administration system. ICT's role in higher education is solicited for improving quality, widening access and enhancing operational efficiency across all functions in higher education sector and to create new dynamics in higher education both at micro and macro levels. Hence, at present the integration of ICT into higher education has become inevitable.

II. Theoretical Background

Change has been happening at an uneven pace in any growth-oriented industry, and the education sector is no exception. In the current age that we live in, technology has become an important component. A tremendous growth in the field of education has made governance in academic communication a very complex task. The 21st century has witnessed rapid advancements in technology which has led to far-reaching developments in the administrative system. Cost-effective technology combined with the flexibility in learning and administrative activities is essential to enhance efficiency. Computers can be used extensively for educational administration. The following are some of the areas where computers can be used for effective educational administration (Ben-Zion Barta et. al. 1995):

- General Administration
- Payroll and Financial Accounting
- Administration of Student Data
- Inventory Management
- Personnel Records Maintenance
- Library System

Rapidly growing computer technology embraced more centers of learning; procedures became necessary to respond to the influx of demands generated by faculty, students and administrators (Maynard J. Bratlien 1989). Information and Communication Technology plays a vital role in supporting powerful, efficient management and administration in the education sector. Technology can be used right from student administration to various resource administrations in an education institution (Christiana Maki 2008). The increasing student population in higher education accelerated the need for ICTs to process, store and retrieve data in a fast, systemic and accurate fashion. The focus of e-administration in higher education is on the creation of an efficient electronic administration by handling existing resources economically. It aims at adding value to the educational sector by simplification of a lot of diversified management and administrative tasks. According to Sanat Kaul (2006), the usage of ICT in higher education institutions starts from the early stages of receiving e-notifications regarding admission, course schedules, and billing procedures and continues till the end of the course including online publication of results. The Action Plan for reforms in the field of higher education revealed that “The potential of Information Communication Technology (ICT) should be fully utilized” (UNESCO 1998).

The main functional areas of e-administration namely, Knowledge administration and Information administration was identified based on the fact that both instructional and managerial are the two main perspectives of administrative tasks in any higher education institutions. Some studies have empirically considered only the knowledge administration that is activities relating to teaching-learning process and hence studies pertaining to Information administration were lacking. Consequently, the model was formed by linking both the functional areas so as to represent a whole model that could be implemented in higher education institutions.

The functional areas that contribute towards e-administration are discussed below along with the constructs that contribute to these functional areas. The study done by Diem Ho (2007) and many others revealed that Knowledge cycle consists of knowledge acquisition, assimilation and development, and evaluation. Based on research studies and literature reviews, it can be concluded that knowledge administration is a cyclic process comprising Knowledge acquisition / Knowledge enhancement, Knowledge delivery and Knowledge evaluation. These include e-learning, e-library, on-line examination, and evaluation and feedback systems as an integral part.

Information administration is considered as another vital part of e-administration. It is required to have continued focus on the contribution of higher education to the national and global economy. This aspect emphasizes the importance of performing educational administration of higher educations with effectiveness and efficiency. Various literature reviews reveal that information administration is one part of overall administration of education institutions which mainly covers general and day-to-day operational activities. Hence, it could be concluded that Information administration cycle includes four major components namely, Student administration, Staff administration, and General administration.

The evolution of higher education in India combined with the need to sustain and be competitive in a global scenario requires decisions to be taken quickly and effectively. This has enhanced the scope and complexity of administration, thus making it necessary to adopt different methods of higher education administration. In this regard, improved levels of deployment of e-administration in higher education institutions are being considered. A theoretical model has been formulated for e-administration.

III. Methodology

The items that contribute towards the two major functional areas namely Knowledge administration and Information administration were carefully identified through extensive literature review and discussions with educational practitioners and experts. A questionnaire was prepared and validated through the pilot study. The reliability and AVE were calculated. The relevant functional areas of e-administration were also profiled together with their mean and standard deviation scores. This was done to examine the average values and the variation. The relationships between the indicators for e-administration were examined. Correlation technique was used for analyzing the association between the indicators of e-administration. Pearson correlation test using SPSS was done to establish this. Finally path diagram constructed using visual PLS is estimated using bootstrapping technique and evaluated using the fit indices.

IV. Analysis And Inferences

The overall mean score for the functional areas of e-administration was 3.55 which revealed that all the functional areas contribute well towards e-administration. It can be observed that knowledge delivery and evaluation has the highest mean value (Table 1.1), and general administration has the lowest mean value. Therefore the sample represents the highest importance assigned to knowledge delivery and evaluation than general administration in technology-based administration in higher education institutions. As revealed by various literature reviews, ICT plays a vital role in the teaching-learning process and the sample also indicates that the application of technology is more visible in knowledge delivery and evaluation than general administrative activities. Though general administration is an important element of e-administration, it is clear from the responses that it is not widely available in practice.

TABLE I. PROFILE OF FUNCTIONAL AREAS OF e-ADMINISTRATION

Functional Areas of e-Administration		
Functional Areas	***Mean Statistic***	***Std. Deviation Statistic***
Knowledge Acquisition and Enhancement	3.92	0.758
Knowledge Delivery and Evaluation	3.97	0.631
Knowledge Administration	3.73	0.722
Student Administration	3.52	0.756
Staff Administration	3.30	0.732
General Administration	2.82	0.531
Information Administration	3.54	0.553

Analysis was done to see whether the indicators for e-administration had a statistically significant relationship with each other. Pearson correlation test revealed that all the constructs correlate positively with each other at 0.01 or 0.05 level of significance.

TABLE II. CORRELATION BETWEEN THE INDICATORS OF E-ADMINISTRATION

	Indicators of e-Administration						
	kae	***kd***	***ka***	***stu***	***stf***	***ga***	***ia***
kae	1						
Kd	0.717 (**)	1					
ka	0.623 (**)	0.687 (**)	1				
stu	0.272 (*)	0.228 (*)	0.332 (**)	1			
stf	0.340 (**)	0.477 (**)	0.478 (**)	0.435 (**)	1		
ga	0.195	0.266 (*)	0.276 (*)	0.591 (**)	0.356 (**)	1	
ia	0.224 (*)	0.311 (**)	0.286 (*)	0.300 (**)	0.320 (**)	0.301 (**)	1

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

The items contributing to knowledge acquisition / enhancement and knowledge delivery were grouped to find the extent to which technology is currently in use. This was done to identify the extent of usage of technology for overall knowledge administration construct (ka). It was inferred from the analysis that 73.3% of the respondents utilized technology for Knowledge acquisition and enhancement, and 75.7% for Knowledge delivery. The analysis revealed that the current technology utilization is 65.1% for overall knowledge administration and 58.5 % for Information administration.

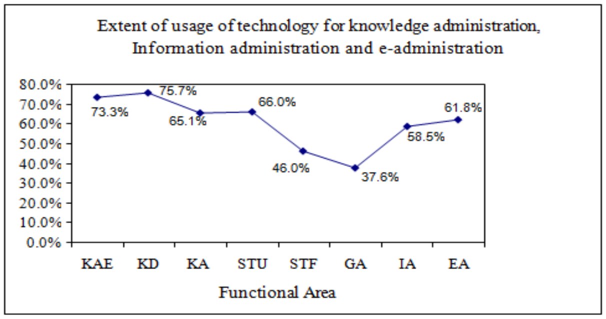
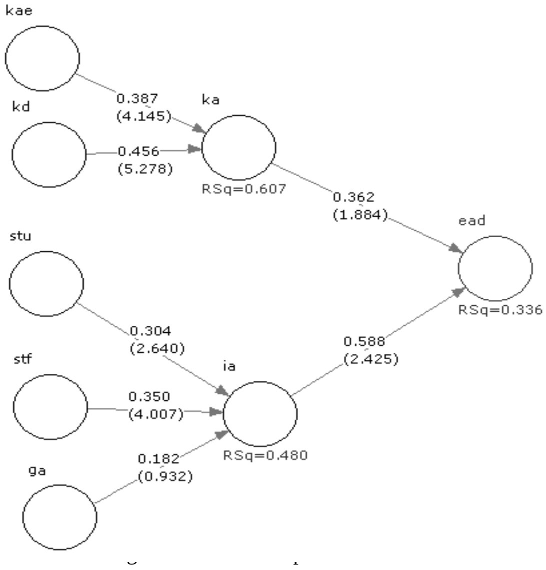


Figure 1. Extent of usage of technology for knowledge administration, information administration and e-administration

V. FINDINGS

- The demographic factors were profiled and their impact on e-administration was studied. It was found that demographic factors do not have a major impact on e-administration in higher education institutions.
- The extent to which technology is being currently used in the various functional areas of e-administration in higher education institutions was analyzed. The usage of technology for Knowledge administration activities was found to be higher than that of Information administration activities
- The functional areas of e-administration and the items constituting them were ranked on the basis of utilization of technology. This revealed that technology was used to the maximum extent in the area of Knowledge delivery and evaluation which is an important indicator for the functional area of knowledge administration.
- Though General administration is an important functional area of Information administration, the utilization of technology was found to be the least among all the functional areas.
- Among the items related to the functional areas of knowledge administration, the highest usage of technology was for Internet browsing by faculty members to supplement book information and the least usage was for the conduct of online tests / online quizzes
- The highest usage of technology in the area of Information administration was found in the area of admission processes through electronic media and the least usage was regarding electronic mode of payment of fees by students.
- All the identified functional areas were found to have an impact on overall e-administration in higher education institutions
- A model relating the various functional areas of knowledge administration and information administration leading to e-administration was estimated and found fit as shown in figure-2 below.



VI. CONCLUSION

This study has identified a comprehensive set of functional areas of e-administration. This study revealed that demographic factors do not have a major impact on e-administration in higher education institutions. It is also evident from this study that integration of ICT into knowledge administration for the teaching–learning process is more in comparison with information administration. Hence, enhancing the usage of ICT on to these functional areas will improve the overall e-administration. This study could serve as a base for education planners and academicians to deploy ICT-based administration in higher education to enhance overall quality of the system.

REFERENCES

[1] Balasubramaniam, Background paper from the Commonwealth of Learning for the UNESCO World Conference on Higher Education**.** Paris, July 6 to 8, 2009.

[2] Ben-Zion Barta, "Information Technology in Educational Management", Chapman and Hall, London, 1995.

[3] Chirstiana Maki, "Information and Communication Technology for Administration and Management for secondary schools in Cyprus", Journal of Online Learning and Teaching Vol. 4 No. 3, 2008.

[4] Diem Ho,"Research, Innovation and Knowledge Management: the ICT Factor", Submitted to UNESCO, July 20, 2007.

[5] Harsh.V. Pant, "Crisis in higher education Universities are in the news more for politics than for research", The Tribune special on-line edition 2009 (http://www.tribuneindia.com).

[6] Maynard J. Bratlien, "Computer Development in Educational Administration", the Journal (Technological Horizons in Education), Vol. 16, 1989.

[7] The Department of Education, Ministry of Human Resource Development, Government of India. Report 2009-2010 (http://www.education.nic.in)

[8] UNESCO World Conference on Higher Education for the Twenty-First Century: Vision and Action, Commission II: Quality of Higher Education, Final Report, Paris: UNESCO, pp.34., 1998.

[9] Sanat Kaul, "Higher Education in India: Seizing the Opportunity", (ICIER Working Paper No. 179), New Delhi: Indian Council for Research on International Economic Relations May 2006. (http://www.icrier.org)

AUTHORS PROFILE

J. Meenakumari

Meenakumari has submitted her doctoral thesis in the integrated topic of Management and IT. She has completed her post graduation and M. Phil in Computer science with distinction. She has fifteen years of teaching experience which includes corporate training and teaching in post graduate courses. Her experience also includes a stint with the Tata group as a center in-charge related to academics and conduct of examinations.

She has been a Keynote speaker and Session chair for the International Conference ICETC 2009 organized by IEEE and IACSIT held at Singapore and also in various international and national conferences and seminars. She has presented papers in various international and national conferences. She won two best paper awards in two international conferences. She has authored research articles in refereed international and national journals. She is a committee member of International Association of Computer Science and Information Technology (IACSIT) and a life member of ISTE. She is also a reviewer for reputed international journals.

Dr. R. Krishnaveni

Dr. R. Krishnaveni is presently Professor in PSGIM, PSG College of Technology, Coimbatore, Tamilnadu. She has twenty five years of teaching and research experience. Her publications include 4 books and 60 research articles in international and national journals. She has been instrumental in organizing National level conferences as well as workshops annually in the area of Business Research. Her popular book includes "Human Resource Development – a Researcher's Perspective" (2008). She is the executive editor of the journal titled "Journal of contemporary research in management ".

Enhanced Architecture of a Web Warehouse based on Quality Evaluation Framework to Incorporate Quality Aspects in Web Warehouse Creation

Umm-e-Mariya Shah
Computer Science Department
COMSATS Institute of Information
Technology Islamabad, Pakistan

Azra Shamim
Computer Science Department
COMSATS Institute of Information
Technology, Islamabad, Pakistan

Madiha Kazmi
Computer Science Department
COMSATS Institute of Information
Technology Islamabad, Pakistan

Abstract—**In the recent years, it has been observed that World Wide Web (www) became a vast source of information explosion about all areas of interest. Relevant information retrieval is difficult from the web space as there is no universal configuration and organization of the web data. Taking the advantage of data warehouse functionality and integrating it with the web to retrieve relevant data is the core concept of web warehouse. It is a repository that store relevant web data for business decision making. The basic function of web warehouse is to collect and store the information for analysis of users. The quality of web warehouse data affects a lot on data analysis. To enhance the quality of decision making different quality dimensions must be incorporated in web warehouse architecture. In this paper enhanced web warehouse architecture is proposed and discussed. The enhancement in the existing architecture is based on the quality evaluation framework. The enhanced architecture adds three layers in existing architecture to insure quality at various phases of web warehouse system creation. The source assessment, query evaluation and data quality layers enhance the quality of data store in web warehouse.**

Keywords-component; Data Warehouse; Web Warehouse; Quality Assessment, Quality Evaluation Framework; Enhanced Web Warehouse Architecture; WWW

I. INTRODUCTION

Due to tremendous advances and achievements in information technology data is being generated at tremendous speed. World Wide Web plays a vital role in information retrieval. World Wide Web has grown to be a universal source and is globally used by individuals and business organizations for information sharing and exchange. Massive amount of data available on the web that is distributed, heterogeneous and semi-structured in nature. Relevant information retrieval is difficult from the web space as there is no universal configuration and organization of the web data [1], [2]. Proper management and retrieval mechanism is required to analyze the information. The distributed and heterogeneous nature of the web data tends to adapt the approach of web warehousing. Currently browsers and search engines are used for information retrieval. Due to lack of knowledge; many search engines may not fully utilize link information [1]. As a result search engines are not able to support such queries or fail to return link information [3]. Web servers can not keep track of the diverse behavior of the client's requests and does not offer the services of web personalization [1]. Therefore an intermediate storage area between the web servers and the clients proves to be a valuable resource [4]. The intermediate repository may not only serve as the storage area but also keeps track of the client's activities and helps in the web personalization [5]. Web warehousing can overcome this problem.

In this research work; the authors enhanced the basic architecture of the Web Warehouse presented in [6] on the basis of the quality evaluation framework discussed in [1]. The authors contribute towards the addition of the quality assessment layer at the time of source selection. A query evaluation layer is embedded with the query processor that may facilitate in query processing. Moreover data quality assessment layer is incorporated between the merge and the load process of the web warehousing system for maintaining high quality of data in web warehouse.

The rest of paper is organized into different sections. Section 2 provides the Literature Review regarding Data Warehouse, Web Warehouse, quality Evaluation Framework. Section 3 consists of the enhanced architecture of the web warehouse based on the quality evaluation framework discussed in section 2. Section 4 presents concluding marks.

II. LITERATERA REVIEW

A. Data Warehouse

Data Warehouse is a central repository that supports executive decision making. According to Hoffer et. al. "A data warehouse (DWH) is an 'informational database' that is maintained separately from an organization's operational database" [1], [7]. "A collection of corporate information, derived directly from operational systems and some external data sources. Its specific purpose is to support business decisions, not business operations" [1], [8]. According to Inmon, "A Data Warehouse is a subject-oriented, integrated, time-variant, non volatile collection of data in support of management decisions" [1], [9], [10], [11], [12]. Different architectures of data warehouse are discussed in [7], [8], [13], [14].

B. Web Warehouse

A Web warehouse is a combination of data warehousing technology and the web technology. According to Mattison, "It is an approach to the building of computer systems which has as its primary functions the identification, cataloguing, retrieval, (possibly) storage, and analysis of information (in the form of data, text, graphics, images, sounds, videos, and other multimedia objects) through the use of Web technology, to help individuals find the information they are looking for and analyse it effectively"[1], [16].

Web warehouse is an architecture comprising of some tools and processes necessary to build up an efficient data warehouse that works on the web and is based on web technologies. Its main functionality is the organization and analysis of stored data and its proper administration. The sources of a web warehouse are the web sites. The web warehouse stores organize and manage the information from web sources and works passively on it. The basic functions of web warehouse include Information Sharing and Intelligent Caching [4]. The contributors of Web Warehouse are Web Technology and Data Warehusing [17]. Different architecture of web warehouse is discussed in [18], [19],[20], [21].

C. Quality Evaluation Framework for a Web Warehousing System

Quality is one of the important factors for the success and survival of any system. To improve quality of a Web Warehouse, Maria et. al. proposed a quality evaluation framework in [1]. This frame work is based on [23], [24],[25] and [26].

Proper evaluation/validation of each phase of web warehouse against certain attributes to measure the quality of the system can be achieved through this framework. It shows categories and dimensions of the quality factors. Further more relevancy of these categories to the phases and sub-phases of a Web Warehouse is also discussed. Quality attributes of discussed in the framework are accessibility, interpretability, usefulness, believability, navigation, efficiency, authority, currency, availability, information-to-noise ratio, popularity, cohesiveness, integrity, reliability, functionality, efficiency, and maintainability.

III. PROPOSED ARCHITECTURE

Enhanced architecture based on the quality evaluation to improve the quality of Web Warehouse System is presented in layered approach in figure 1. The main components of the proposed architecture are described as follow:

A. Web Information Sources

The web warehouse is constructed over a heterogeneous, distributed and semi-structured web space. The data is gathered from different web sites. It undergoes the process of transformation and integration and stored in a web warehouse. The web information sources are defined at the time of making web warehouse design specification. However new web information sources can be included. It involves creation of data source view in a web warehouse along with the creation of the respective view manager. The newly added data source is then connected with the respective monitor and wrapper.

B. Source Evaluation Layer

This layer assesses the origin of the data that is selected for data extraction. Assessment is on the basis of some quality dimensions i.e. source currency, relevancy, availability, information-to-noise ratio, authority, popularity and cohesiveness. It measures the number of broken links on a web page, proportion of the useful information, prestige of the data source, relevancy of the major topics in a web page and the number of citations by other web pages. The evaluation ensures that there exist up-to-date contents in the selected source that are compatible to the user's query. The source evaluation layer then filters out the data belonging to those sources only that satisfies the evaluation criteria.

C. Monitor

This component is connected to the underlying information sources. Each data source has its monitor. It polls the web information sources periodically to detect any changes arise in them. Polling is done by comparing the snapshots and obtaining the base data changes. It collects the changes and notifies the modifications to the integrator component.

D. Wrapper

This component deals with the data extraction. It accepts the query from query processor and mines results from the underlying information sources. The result is then transformed into a specified format of a web warehousing system.

E. Integrator

This component maintains consistency between the web warehousing system and the underlying information resources. Any updated information is sent to the integrator by the monitor. Integrator integrates the information and sends the modifications to the respective view manager.

F. View Manager

It keeps the consistency between the views in a web warehouse and views of the data sources. Every view in a web warehouse has its own view manager to perform all necessary actions. Whenever a change is detected in the underlying information resource it is transferred by the monitor to the integrator. Integrator then sends the modification to the relevant view manager. The view manager then updates the relevant web view.

G. Query Processor and Evaluator

Whenever a query is initiated by a user it is transferred to the query processor and evaluator. Query evaluator assesses the query on the basis of certain quality dimensions like semantics, performance, time behavior and optimization. It ensures that the system must timely respond to a directed task and makes certain the clarity of its meaning, structure and language rules. Query processor then executes the refined query. It translates the query from the high level language to the low level language. It consults the meta data repository and directs the query to the appropriate data source.

H. Merge Process

View managers when perform actions the results are passed to the merge process. The merge process combines and sorts these results according to the user's query sequence.

I. Data Quality Assessment Layer

This layer evaluates the data in terms of certain quality dimensions. Extracted data may be erroneous and contains some inconsistencies.

The data quality assessment layer ensures that the data must be correct, comprehensive precise and stable. Thus high quality data will become available to load in the web warehouse that leads to a quality decision. This layer performs some quality control techniques as described below:

- Data Auditing and Standardization - Typically, the data in data stores and databases is inconsistent and lacks conformity. Data auditing ensures the precision and accuracy of data at the source [27]. It evaluates the data (in the source database) against a set of business rules to perform validation checks. It provides frequencies of data fields and identifies the outliers and the range of value for each attribute. The business and cleansing rules are identified in the data auditing process. The business rules may be determined by using data mining techniques which are used to uncover the patterns in the data. Outlier data is then modified as required.
- Data Linking and Consolidation - The data coming from multiple sources may be inconsistent and redundant. Data linking identifies the records that represent the same values of an entity and links them. In the consolidation process elements of matching records are combined into a complete record [27].
- Information Stewardship - The validity of information can be obtained if automated routines and business rules are implemented but they do not help for information accuracy. People and experts are needed to assure the accuracy [28]. Stewardship is “the willingness to be accountable for the well-being of the larger organization by operating in service of, rather than in control of those around us” [29]. Information stewardship is “the willingness to be accountable for a set of business information for the well-being of the larger organization by operating in service, rather than in control of those around us” [29].
- Data Cleaning - “Data cleansing is a process of identifying and removing errors or inconsistencies from the data in order to improve the quality” [30]. Data quality problems exist in each case whether data has single or multiple sources.

J. Load Process

The load process loads the high quality data to the web warehouse. This data is the result of the user’s query response.

K. Web Warehouse

It is the final destination where the results are stored. Web warehouse has following components.

- Web Marts - These are designed separately for a particular department. Web marts are the subsets of a web warehouse and contains the data satisfying that particular department needs. In this way web marts help to decrease the query response time of the end user.
- Meta Data Repository - Meta data means the data about data. It is a storehouse where all the necessary information regarding extracted data is stored. It is consulted by various processes in performing their tasks. Like wrappers consult the meta data for the relevant data sources before starting the extraction process. Query processor uses the meta data repository to find the appropriate data source for the query execution phenomenon.
- Meta Data Manager - It supports the maintenance of the meta data repository. The data management and manipulation of meta data repository is handled by it.
- Web Manager - Web sites that are selected as the information sources for a web warehouse are managed by the web manager. It makes decision for the addition and deletion of the data sources. It also monitors the performance of the view managers.

L. Presentation Layer

This layer provides interface to the user. It interacts with the web warehouse and extracts the information. The data that is extracted from a web warehouse is analyzed via various tools and helps in decision making.

IV. CONCLUSION

During the web warehouse creation phase the output of one phase becomes the input of the next phase, so quality assessment is most important at various stages of the web warehouse to get a successful web warehousing system. Keeping in view the quality factor an enhanced architecture of web warehouse is proposed and discussed in this paper. The enhanced architecture increases the quality of web warehouse system by introducing source assessment, query evaluation and data quality layer. The basic architecture of a web warehouse is enhanced by embedding the source assessment layer, query evaluation layer and data quality layer. Data quality layer ensures the quality of data before loading it into warehouse. Source assessment layer is responsible for checking validity, relevancy and other quality attributes of web sources. Query evaluation layer facilitate in query processing.

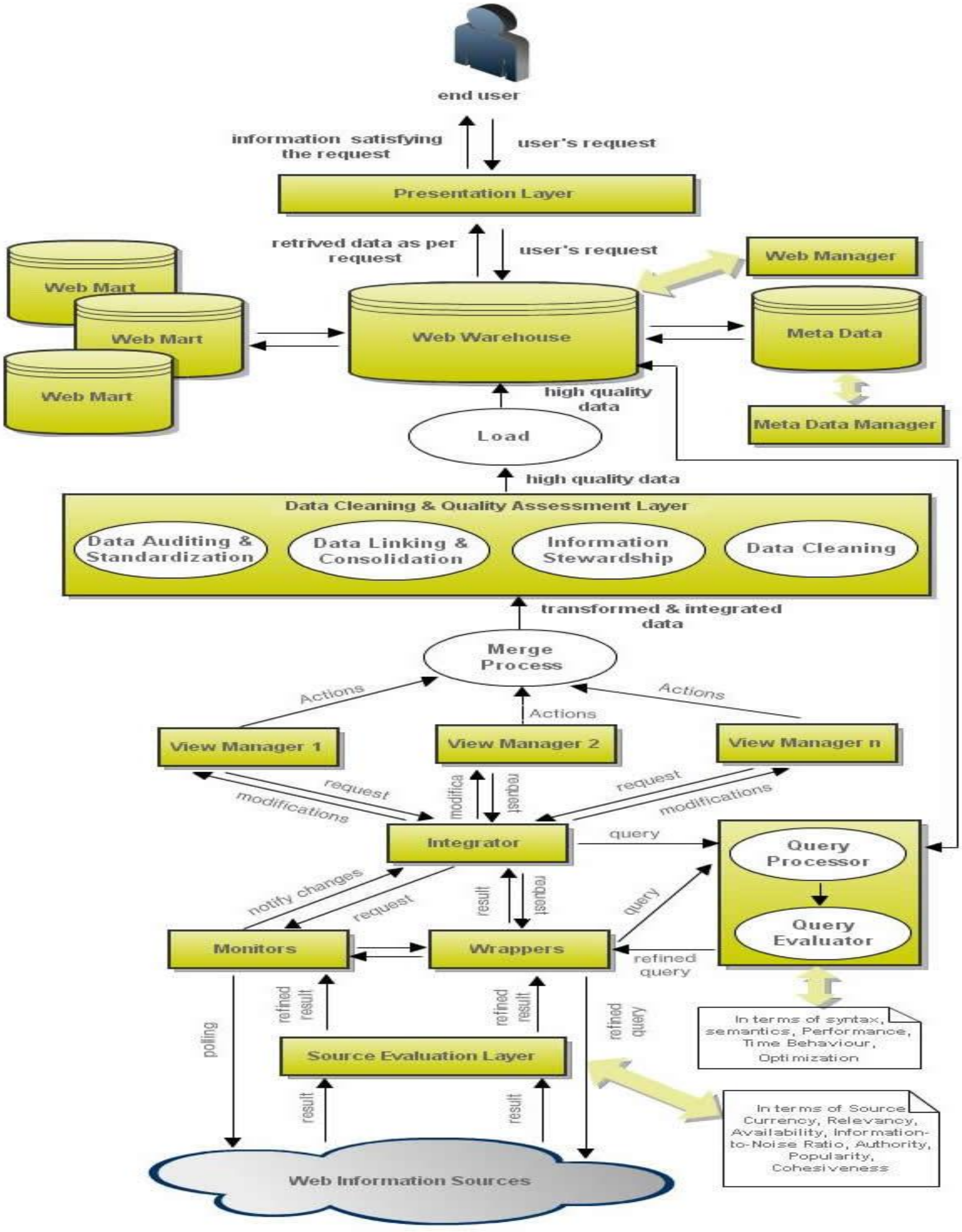


Fig. 1 Enhanced Architecture of a Web Warehouse

References

[1] Umm-e-Mariya Shah, Maqbool Uddin Shaikh, Azra Shamim , Yasir Mehmood, Proposed Quality Evaluation Framework to Incorporate Quality Aspects in Web Warehouse Creation, Journal of Computing, Volume 3, Issue 4, May 2011.

[2] J. Dyche, "e-Data: turning data into information with data warehousing," Addison-Wesley, Reading, MA, 2000

[3] Sourav S. Bhowmick, Wee-Keong Ng, and Ee-Peng Lim, "Information Coupling in Web Database," Springer-Verlag Berlin Heidelberg, pp. 92-106, 1998.

[4] Kai Cheng, Yahiko Kambayashi, Seok Tae Lee and Mukesh Mohania, “Functions of a Web Warehouse,” International Conference on IEEE Digital Libraries: Research and Practice, Kyoto, 2000.

[5] Masahiro Hori, Goh Kondoh, Kohichi Ono, Shin ichi Hirose, and Sandeep Singhal, Annotation-based Web Content Transcoding, http://www9.org/w9cdrom/index.html, Accessed Date: 05 Dec, 2009.

[6] Saif ur Rehman, Maqbool Uddin Shaikh, " “Web Warehouse: Towards Efficient Distributed Business Management”, In proceedings of IEEE International Multi-Topic Conference 2008 (INMIC-2008)

[7] Jeffrey A. Hoffer, Mary B. Prescott, Fred R. McFadden, Modern database management, Sixth Edition, Pearson Education Publishers, Singapore

[8] Thomas Connolly, Carolyn Begg, “Database Systems: A Practical Approach to Design, Implementation and Management,” 4th Edition, Addison-Wesley, 2003

[9] William Inmon, Building the Data Warehouse, 2nd Edition, New York: Wiley publisher. Inc, 1996

[10] Rizwana Irfan, Azra Shamim, Madiha Kazmi, Framework for Case Based Object Oriented Expert Warehouse to Enhance Knowledge Management Process for Executive Decision Making, Journal of Computing, Volume 3, Issue 4, May 2011.

[11] Atika Qazi, Azra Shamim, Rubina Adnan, Farooq Azam, A Distributed Data Warehouse Architecture with Fair Query Execution Scheme, ICMLC, 2011

[12] Saif Ur Rehman Malik, Azra Shamim, Zanib Bibi, Sajid Ullah Khan, Shabir Ahmad Gorsi, A Framework for ETL Workflow Management for Efficient Business Decision-Making, ICSCT 2010

[13] Daniel L. Moody, Mark A.R. Kortink, “From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design”, In Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'2000) June 5-6, 2000, Stockholm, Sweden

[14] Mohammad Rifaie, Erwin J. Blas, Abdel Rahman M. Muhsen, Terrance T. H. Mok, Keivan Kianmehr, Reda Alhajj, Mick J. Ridley, “Data warehouse Architecture for GIS Applications”, In Proceedings of the 10th International Conference on Information Integration and Web-based Applications & Services (iiWAS '08) , November 2008, Linz, Austria

[15] Data Warehousing and OLAP, www.cs.uh.edu/~ceick/6340/dw-olap.ppt

[16] R. Mattison, “Web warehousing and knowledge management,” 1st Edition, New York: McGraw-Hill School Education Group, 1999.

[17] Xin Tan, David C. Yen and Xiang Fang, “Web Warehousing: Web technology meets data warehousing,” Science Direct Technology in Society, 2003.

[18] Kai Cheng, Yahiko Kambayashi, Seok Tae Lee and Mukesh Mohania, “Functions of a Web Warehouse,” International Conference on IEEE Digital Libraries: Research and Practice, Kyoto, 2000.

[19] Lean Yu, Wei Huang, Shouyang Wang, Kin Keung Lai, “Web warehouse – a new web information fusion tool for web mining,” Elsevier, information fusion, science direct, 2006.

[20] Web Data Warehousing, “DVS, web data warehousing”, Available: http://www.dvs.tu-darmstadt.de/research/webdataware/, Access Date: 5 December 2008

[21] Yan Zhang and Xiangdong Qin, “Effectively Maintaining Single View Consistency in Web Warehouses,” CIT The Fifth International Conference on Computer and Information Technology, IEEE computer Society, pp 199-205, 2005.

[22] Panos Vassiliadis, “Data Warehouse Modeling and Quality Issues,” National Technical University of Athens Zographou, Athens, GREECE, 2000.

[23] A Framework for Assessing Database Quality, http://osm7.cs.byu.edu/ER97/workshop4/jh.html, Accessed Date: May 16, 2010.

[24] Informatik V , Matthias Jarke , Matthias Jarke , Lehrstuhl Fur Informatik V , Yannis Vassiliou , Yannis Vassiliou Asdasda, “Data Warehouse Quality: A Review of the DWQ Project”, In Proceedings of the 2nd Conference on Information Quality, Massachusetts Institute of Technology, Cambridge, 1997.

[25] Xiaolan Zhu, Susan Gauch, “Incorporating quality metrics in centralized/distributed information retrieval on the World Wide Web”, In Proceedings of the 23rd annual international conference on Research and development in information retrieval, ACM SIGIR, 2000.

[26] Shirlee-ann Knight, Janice Burn, “Developing a Framework for Assessing Information Quality on the World Wide Web”, The World Wide Web. Informing Science Journal, 2005.

[27] M. Pamela Neely “Data Quality Tools for Data Warehousing – A Small Sample Survey,” Center for Technology in Government University at Albany / SUNY, 1998.

[28] Larry P. English, “Information Stewardship: Accountability for Information Quality,” Information Impact International, Inc, 2006.

[29] Peter Block, “Stewardship: Choosing Service over Self-Interest,” San Francisco: Berett-Koehler, 1993.

[30] Erhard Rahm, Hong Hai Do, “Data Cleaning: Problems and Current Approaches,” Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2000.

AUTHORS PROFILE

Umm-e-Mariya Shah is a student of MS(CS) in COMSATS Institute of Information Technology. In addition she is an IT Consultant in Sustainalbe Development Policy Institute, Islamabad. Pakistan. Also she is a visiting lecture in SKANS School of Accountancy, Rawalpindi, Pakistan.

Azra Shamim is working as a research associate in COMSATS Institute of Information Technology, Islamabad, Pakistan. She received her MS(CS) degree from COMSATS Institute of Information Technology, Islamabad, Pakistan.

Madiha Kazmi is working as a Lecture in COMSATS Institute of Information Technology, Islamabad, Pakistan. She received her MS(CS) degree from from National University of Science and Technology, Islamabad, Pakistan.

Detection of Reliable Software Using SPRT

Dr. R.Satya Prasad
Dept. of CS & Engg.
Acharya Nagrjuna University
Guntur, Andhrapradesh,
India

N. Supriya
Dept. of Computer Science, Adikavi
Nannaya University Rajahmedndry,
Andhrapradesh,
India

G.Krishna Mohan
Dept. of Computer Science
P.B.Siddhartha College, Vijayawada,
Andhrapradesh,
India

Abstract— **In Classical Hypothesis testing volumes of data is to be collected and then the conclusions are drawn which may take more time. But, Sequential Analysis of statistical science could be adopted in order to decide upon the reliable / unreliable of the developed software very quickly. The procedure adopted for this is, Sequential Probability Ratio Test (SPRT). In the present paper, we have proposed the performance of SPRT on Time domain data using exponential imperfect debugging model and analyzed the results by applying on 5 data sets. The parameters are estimated by using Maximum Likelihood Estimation.**

***Keywords-** Exponential imperfect debugging; Sequential Probability Ratio Test; Maximum Likelihood Estimation; Decision lines; Software Reliability; Software failure data.*

I. INTRODUCTION

Wald's procedure is particularly relevant if the data is collected sequentially. Sequential Analysis is different from that of Classical Hypothesis Testing where the number of cases tested or collected, is fixed at the beginning of the experiment. In Classical Hypothesis Testing, the data collection is executed without analysis and consideration of the data. After all the data is collected the analysis is done, conclusions are drawn. However, in Sequential Analysis every case is analyzed directly after being collected, the data collected up to that moment is then compared with certain threshold values, incorporating the new information obtained from the freshly collected case. This approach allows one to draw conclusions during the data collection, and a final conclusion can possibly be reached at a much earlier stage as is the case in Classical Hypothesis Testing. The advantages of Sequential Analysis is easily seen. As data collection can be terminated after fewer cases and decisions taken earlier, the savings in terms of human life and misery, and financial savings, might be considerable.

In the analysis of software failure data, we often deal with either Time Between Failures or failure count in a given time interval. If it is further assumed that the average number of recorded failures in a given time interval is directly proportional to the length of the interval and the random number of failure occurrences in the interval is explained by a Poisson process, then we know that the probability equation of the stochastic process representing the failure occurrences is given by a homogeneous poisson process with the expression

$$P\left[N(t)=n\right]=\frac{e^{-\lambda t}(\lambda t)^{n}}{n!} \qquad (1.1)$$

Stieber [5] observes that if classical testing strategies are used, the application of software reliability growth models may be difficult and reliability predictions can be misleading. However, he observes that statistical methods can be successfully applied to the failure data. He demonstrated his observation by applying the well-known sequential probability ratio test of Wald [4] for a software failure data to detect unreliable software components and compare the reliability of different software versions. In this paper we consider popular SRGM Exponential imperfect debugging model and adopt the principle of Stieber in detecting unreliable software components in order to accept or reject the developed software. The theory proposed by Stieber is presented in Section 2 for a ready reference. The extension of this theory to the SRGM – Exponential imperfect debugging is presented in Section 3. The Maximum Likelihood parameter estimation method is presented in Section 4. and Application of the decision rule to detect unreliable software components with respect to the proposed SRGM is given in Section 5.

II. WALD'S SEQUENTIAL TEST FOR A POISSON PROCESS

The sequential probability ratio test was developed by A.Wald at Columbia University in 1943. Due to its usefulness in development work on military and naval equipment it was classified as ‘Restricted’ by the Espionage Act (Wald, 1947). A big advantage of sequential tests is that they require fewer observations (time) on the average than fixed sample size tests. SPRTs are widely used for statistical quality control in manufacturing processes. An SPRT for homogeneous Poisson processes is described below.

Let $\{N(t), t \geq 0\}$ be a homogeneous Poisson process with rate ‘λ’. In our case, N(t) = number of failures up to time ‘ t’ and ‘λ’ is the failure rate. Suppose if we put a system on test (for example a software system, where testing is done according to a usage profile and no faults are corrected) and that we want to estimate its failure rate ‘λ’. We cannot expect to estimate ‘λ’ precisely. But if we want to reject the system with a high probability, our data suggest that the failure rate is larger than λ_1 and accept it with a high probability, if it’s smaller than λ_0. As always with statistical tests, there is some risk to get the wrong answers. So we have to specify two (small) numbers ‘α’ and ‘β’, where ‘α’ is the probability of falsely rejecting the system. That is rejecting the system even if $\lambda \leqslant \lambda_0$. This is the "producer’s" risk. β is the probability of falsely accepting the system .That is accepting the system even if $\lambda \geqslant \lambda_1$. This is the "consumer's" risk. With specified choices of λ_0 and λ_1 such that $0 < \lambda_0 < \lambda_1$, the probability of finding N(t) failures in the time span (0,t) with λ_1, λ_0 as the failure rates are respectively given by

$$P_1 = \frac{e^{-\lambda_1 t}\left[\lambda_1 t\right]^{N(t)}}{N(t)!} \qquad (2.1)$$

$$P_0 = \frac{e^{-\lambda_0 t}\left[\lambda_0 t\right]^{N(t)}}{N(t)!} \qquad (2.2)$$

The ratio $\frac{p_1}{p_0}$ at any time 't' is considered as a measure of deciding the truth towards λ_0 or λ_1, given a sequence of time instants say $t_1 < t_2 < t_3 < \ldots\ldots < t_K$ and the corresponding realizations $N(t_1), N(t_2), \ldots\ldots N(t_K)$ of N(t). Simplification of $\frac{p_1}{p_0}$ gives

$$\frac{P_1}{P_0} = \exp(\lambda_0 - \lambda_1)t + \left(\frac{\lambda_1}{\lambda_0}\right)^{N(t)}$$

The decision rule of SPRT is to decide in favor of λ_1, in favor of λ_0 or to continue by observing the number of failures at a later time than 't' according as $\frac{p_1}{p_0}$ is greater than or equal to a constant say A, less than or equal to a constant say B or in between the constants A and B. That is, we decide the given software product as unreliable, reliable or continue [3] the test process with one more observation in failure data, according as

$$\frac{P_1}{P_0} \geq A \qquad (2.3)$$

$$\frac{P_1}{P_0} \leq B \qquad (2.4)$$

$$B < \frac{P_1}{P_0} < A \qquad (2.5)$$

The approximate values of the constants A and B are taken as $A \cong \frac{1-\beta}{\alpha}$, $B \cong \frac{\beta}{1-\alpha}$

Where 'α' and 'β' are the risk probabilities as defined earlier. A simplified version of the above decision processes is to reject the system as unreliable if N(t) falls for the first time above the line

$$N_U(t) = a.t + b_2 \qquad (2.6)$$

To accept the system to be reliable if N(t) falls for the first time below the line

$$N_L(t) = a.t - b_1 \qquad (2.7)$$

To continue the test with one more observation on (t, N(t)) as the random graph of [t, N(t)] is between the two linear boundaries given by equations (2.6) and (2.7) where

$$a = \frac{\lambda_1 - \lambda_0}{\log\left(\frac{\lambda_1}{\lambda_0}\right)} \qquad (2.8)$$

$$b_1 = \frac{\log\left[\frac{1-\alpha}{\beta}\right]}{\log\left(\frac{\lambda_1}{\lambda_0}\right)} \qquad (2.9)$$

$$b_2 = \frac{\log\left[\frac{1-\beta}{\alpha}\right]}{\log\left(\frac{\lambda_1}{\lambda_0}\right)} \qquad (2.10)$$

The parameters α, β, λ_0 and λ_1 can be chosen in several ways. One way suggested by Stieber is

$$\lambda_0 = \frac{\lambda.\log(q)}{q-1}, \lambda_1 = q\frac{\lambda.\log(q)}{q-1}, \text{ where } q = \frac{\lambda_1}{\lambda_0}$$

If λ_0 and λ_1 are chosen in this way, the slope of $N_U(t)$ and $N_L(t)$ equals λ. The other two ways of choosing λ_0 and λ_1 are from past projects (for a comparison of the projects) and from part of the data to compare the reliability of different functional areas (components).

III. Sequential Test for Software Reliability Growth Models

In Section 2, for the Poisson process we know that the expected value of N(t) = λt called the average number of failures experienced in time 't' .This is also called the mean value function of the Poisson process. On the other hand if we consider a Poisson process with a general function (not necessarily linear) m(t) as its mean value function the probability equation of a such a process is

$$P\left[N(t) = Y\right] = \frac{\left[m(t)\right]^y}{y!}.e^{-m(t)}, y = 0, 1, 2, ----$$

Depending on the forms of m(t) we get various Poisson processes called NHPP. For our model the mean value function is given as $m(t) = \frac{a}{1-c}\left(1 - e^{-(1-c)bt}\right)$ where $a > 0, b > 0$

We may write

$$P_1 = \frac{e^{-m_1(t)}.\left[m_1(t)\right]^{N(t)}}{N(t)!}$$

$$P_0 = \frac{e^{-m_0(t)}.\left[m_0(t)\right]^{N(t)}}{N(t)!}$$

Where, $m_1(t), m_0(t)$ are values of the mean value function at specified sets of its parameters indicating reliable software and unreliable software respectively. Let P_0, P_1 be values of the NHPP at two specifications of b say b_0, b_1 where $(b_0 < b_1)$ respectively. It can be shown that for our model

$m(t)$ at b_1 is greater than that at b_0. Symbolically $m_0(t) < m_1(t)$. Then the SPRT procedure is as follows:

Accept the system to be reliable $\frac{P_1}{P_0} \le B$

$$\text{i.e., } \frac{e^{-m_1(t)}.\left[m_1(t)\right]^{N(t)}}{e^{-m_0(t)}.\left[m_0(t)\right]^{N(t)}} \le B$$

$$\text{i.e., } N(t) \le \frac{\log\left(\frac{\beta}{1-\alpha}\right) + m_1(t) - m_0(t)}{\log m_1(t) - \log m_0(t)} \qquad (3.1)$$

Decide the system to be unreliable and reject if $\frac{P_1}{P_0} \ge A$

$$\text{i.e., } N(t) \ge \frac{\log\left(\frac{1-\beta}{\alpha}\right) + m_1(t) - m_0(t)}{\log m_1(t) - \log m_0(t)} \, (3.2)$$

Continue the test procedure as long as

$$\frac{\log\left(\frac{\beta}{1-\alpha}\right) + m_1(t) - m_0(t)}{\log m_1(t) - \log m_0(t)} < N(t) < \frac{\log\left(\frac{1-\beta}{\alpha}\right) + m_1(t) - m_0(t)}{\log m_1(t) - \log m_0(t)}$$

(3.3)

Substituting the appropriate expressions of the respective mean value function – m(t) of the considered model we get the respective decision rules and are given in followings lines

Acceptance region:

$$N(t) \le \frac{\log\left(\frac{\beta}{1-\alpha}\right) + \frac{a}{1-c}\left(e^{-(1-c)b_0 t} - e^{-(1-c)b_1 t}\right)}{\log\left(\frac{1-e^{-(1-c)b_1 t}}{1-e^{-(1-c)b_0 t}}\right)} \quad (3.4)$$

Rejection region:

$$N(t) \ge \frac{\log\left(\frac{1-\beta}{\alpha}\right) + \frac{a}{1-c}\left(e^{-(1-c)b_0 t} - e^{-(1-c)b_1 t}\right)}{\log\left(\frac{1-e^{-(1-c)b_1 t}}{1-e^{-(1-c)b_0 t}}\right)} \quad (3.5)$$

Continuation region:

$$\frac{\log\left(\frac{\beta}{1-\alpha}\right) + \frac{a}{1-c}\left(e^{-(1-c)b_0 t} - e^{-(1-c)b_1 t}\right)}{\log\left(\frac{1-e^{-(1-c)b_1 t}}{1-e^{-(1-c)b_0 t}}\right)} < N(t) < \frac{\log\left(\frac{1-\beta}{\alpha}\right) + \frac{a}{1-c}\left(e^{-(1-c)b_0 t} - e^{-(1-c)b_1 t}\right)}{\log\left(\frac{1-e^{-(1-c)b_1 t}}{1-e^{-(1-c)b_0 t}}\right)}$$

(3.6)

It may be noted that in the above model the decision rules are exclusively based on the strength of the sequential procedure (α,β) and the values of the respective mean value functions namely, $m_0(t), m_1(t)$. If the mean value function is linear in ‘t’ passing through origin, that is, m(t) = λt the decision rules become decision lines as described by Stieber (1997). In that case equations (3.1), (3.2) , (3.3) can be regarded as generalizations to the decision procedure of Stieber (1997). The applications of these results for live software failure data are presented with analysis in Section 5.

IV. ML (Maximum Likelihood) Parameter Estimation

The idea behind maximum likelihood parameter estimation is to determine the parameters that maximize the probability (likelihood) of the sample data. The method of maximum likelihood is considered to be more robust (with some exceptions) and yields estimators with good statistical properties. In other words, MLE methods are versatile in their approach and can be applied to many models and also to different types of data. Although the methodology for maximum likelihood estimation is simple, the implementation is mathematically complex. Using today's computer power, however, mathematical complexity is not a big obstacle. If we conduct an experiment and obtain N independent observations, $t_1, t_2, \ldots, t_N$. Then the likelihood function is given by[9] the following product:

$$L(t_1, t_2, \ldots, t_N \mid \theta_1, \theta_2, \ldots, \theta_k) = L = \prod_{i=1}^{N} f(t_i; \theta_1, \theta_2, \ldots, \theta_k)$$

Likely hood function by using λ(t) is: L = $\prod_{i=1}^{n} \lambda(t_i)$

The logarithmic likelihood function is given by: $\Lambda = \ln L = \sum_{i=1}^{N} \ln f(t_i; \theta_1, \theta_2, \ldots, \theta_k)$

Log L = log ($\prod_{i=1}^{n} \lambda(t_i)$); which can be written as $\sum_{i=1}^{n} \log[\lambda(t_i)] - m(t_n)$

The maximum likelihood estimators (MLE) of $\theta_1, \theta_2, \ldots, \theta_k$ are obtained by maximizing L or Λ, where Λ is ln L . By maximizing Λ, which is much easier to work with than L, the maximum likelihood estimators (MLE) of $\theta_1, \theta_2, \ldots, \theta_k$ are the simultaneous solutions of k equations such that: $\frac{\partial(\Lambda)}{\partial\theta_j} = 0$, j=1,2,…,k

The parameters ‘a’ and ‘b’ are estimated using iterative Newton Raphson Method, which is given as

$$x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)}$$

For the present model of Exponential imperfect debugging at c=0.05, the parameters are estimated from [10].

V. SPRT Analysis of Data Sets

We see that the developed SPRT methodology is for a software failure data which is of the form [t, N(t)] where N(t) is the failure number of software system or its sub system in ‘t’ units of time. In this section, we evaluate the decision rules based on the considered mean value function for Five different data sets of the above form, borrowed from [2][7][8] with the assumption of c=0.05. Based on the estimates of the parameter ‘b’ in each mean value function, we have chosen the specifications of $b_0 = b - \delta$, $b_1 = b + \delta$ equidistant on either side of estimate of b obtained through a data set to apply SPRT such that $b_0 < b < b_1$. Assuming the value of $\delta = 0.0025$, the choices are given in the following table.

TABLE I. ESTIMATES OF a,b & SPECIFICATIONS OF b_0, b_1

Data Set	Estimate of 'a'	Estimate of 'b'	b_0	b_1
DS 1 [7]	31.738136	0.003253	0.000753	0.005753
DS 2 [2]	24.182003	0.003091	0.000591	0.005591
DS 3 [2]	22.286839	0.003627	0.001127	0.006127
DS 4 [2]	32.293828	0.006095	0.003595	0.008595
DS 5 [8]	30.276648	0.020823	0.018323	0.023323

Using the selected b_0, b_1 and subsequently the $m_0(t), m_1(t)$ for the model, we calculated the decision rules given by Equations 3.1, 3.2, sequentially at each 't' of the data sets taking the strength (α, β) as (0.05, 0.2). These are presented for the model in Table II.

TABLE II. SPRT ANALYSIS FOR 5 DATA SETS

Data Set	T	N(t)	R.H.S of equation (5.3.10) Acceptance region (≤)	R.H.S of Equation (5.3.11) Rejection Region(≥)	Decision
DS 1	30.02	1	1.419789011	3.625727114	***Accept***
DS 2	5.5	1	-0.404541835	1.53395301	***Reject***
	7.33	2	-0.309766769	1.632482881	
DS 3	10	1	-0.288370452	2.305609383	***Reject***
	19	2	0.258176419	2.885230322	
	32	3	1.012360023	3.688059978	
	43	4	1.619570979	4.337249402	
	58	5	2.404532289	5.180679581	
	70	6	2.998756638	5.822712830	
DS 4	9	1	-0.210017229	4.882283497	***Accept***
	21	2	1.757136003	7.020723877	
	32	3	3.432506995	8.859613125	
DS 5	0.5	1	-6.177485603	11.85991989	***Continue***
	1.7	2	-5.518244758	12.73424974	
	4.5	3	-4.051626961	14.71612777	
	7.2	4	-2.728688612	16.55426597	
	10	5	-1.446988791	18.39005947	
	13	6	-0.17086726	20.28319498	
	14.8	7	0.548597293	21.38490073	
	15.7	8	0.89573932	21.926644	
	17.1	9	1.419425808	22.75775024	
	20.6	10	2.64443639	24.77711192	
	24	11	3.724217517	26.66533205	
	25.2	12	4.08045555	27.31591129	
	26.1	13	4.339358286	27.79874025	
	27.8	14	4.809408109	28.69931804	
	29.2	15	5.178228348	29.43026	
	31.9	16	5.844193721	30.81447884	
	35.1	17	6.558795727	32.4155083	
	37.6	18	7.06283825	33.63975553	
	39.6	19	7.432965441	34.60421986	
	44.1	20	8.162162302	36.73248646	
	47.6	21	8.634040324	38.35477082	
	52.8	22	9.188544561	40.72597288	
	60	23	9.681798744	43.96479429	
	70.7	24	9.861407038	48.78058702	

From the above table we see that a decision either to accept or reject the system is reached much in advance of the last time instant of the data(the testing time).The following consolidated table reveals the iterations required to come to a decision about the software of each data set.

VI. CONCLUSION

The table II shows that The Exponential imperfect debugging model as exemplified for 5 Data Sets indicate that the model is performing well in arriving at a decision. Out of 5 Data Sets, the procedure applied on the model has given a decision of rejection for 2, acceptance for 2 and continue for 1 at various time instant of the data as follows. DS1, DS4 are accepted at 1st and 3rd instant of time respectively, DS2, DS3 are rejected at 2nd and 6th instant of time respectively. DS5 is continuing. Therefore, we may conclude that, applying SPRT on data sets we can come to an early conclusion of reliable / unreliable of software.

REFERENCES

[1] Goel, A.L and Okumoto, K. "A Time Dependent Error Detection Rate Model For Software Reliability And Other Performance Measures", IEEE Transactions on Reliability, vol.R-28, pp.206-211, 1979.

[2] Pham. H., "System software reliability", Springer. 2006.

[3] Satya Prasad, R., "Half logistic Software reliability growth model ", 2007, Ph.D Thesis of ANU, India.

[4] Wald. A., "Sequential Analysis", John Wiley and Son, Inc, New York. 1947.

[5] Stieber, H.A. "Statistical Quality Control: How To Detect Unreliable Software Components", Proceedings the 8th International Symposium on Software Reliability Engineering, 8-12. 1997.

[6] Wood, A. "Predicting Software Reliability", IEEE Computer, 2253-2264. 1996.

[7] Xie, M., Goh. T.N., Ranjan.P., "Some effective control chart procedures for reliability monitoring" -Reliability engineering and System Safety 77 143 -150¸ 2002.

[8] Michael. R. Lyu, "The hand book of software reliability engineering", McGrawHill & IEEE Computer Society press.

[9] Pham. H., 2006. "System software reliability", Springer.

[10] Satya Prasad. R, Supriya. N and Krishna Mohan. G, "spc for software reliability: imperfect software debugging model". International Journal of Computer Science Issues. Vol.8, Issue.3, No.2.

AUTHORS PROFILE

Dr. R. Satya Prasad received Ph.D. degree in Computer Science in the faculty of Engineering in 2007 from Acharya Nagarjuna University, Andhra Pradesh. He received gold medal from Acharya Nagarjuna University for his outstanding performance in Masters Degree. He is currently working as Associate Professor and H.O.D, in the Department of Computer Science & Engineering, Acharya Nagarjuna University. His current research is focused on Software Engineering. He has published several papers in National & International Journals.

Mrs. N.Supriya Working as Academic Consultant, in the department of Computer Science, Adikavi Nannaya university, Rajahmundry. She is at present pursuing Ph.D at Acharya Nagarjuna University. Her research interests lies in Software Engineering and Data Mining.

Mr. G. Krishna Mohan is working as a Reader in the Department of Computer Science, P.B.Siddhartha College, Vijayawada. He obtained his M.C.A degree from Acharya Nagarjuna University in 2000, M.Tech from JNTU, Kakinada, M.Phil from Madurai Kamaraj University and pursuing Ph.D at Acharya Nagarjuna University. His research interests lies in Data Mining and Software Engineering.

Design of an Intelligent Combat Robot for war fields

Dr.S.Bhargavi
Electronics and Communication Engineering
S.J.C.I.T
Chikballapur, Karnataka, India

S.Manjunath
Electronics and Communication Engineering
S.J.C.I.T
Chikballapur, Karnataka, India

Abstract—**The objective of this paper is to minimize human casualties in terrorist attack such as 26/11. The combat robot [1] has been designed to tackle such a cruel terror attacks. This robot is radio operated, self- powered, and has all the controls like a normal car. A wireless camera has been installed on it, so that it can monitor enemy remotely when required. It can silently enter into enemy area and send us all the information through its' tiny Camera eyes. This spy robot can be used in star hotels, shopping malls, jewellary show rooms, etc where there can be threat from intruders or terrorists. Since human life is always precious, these robots are the replacement of fighters against terrorist in war areas.**

Keywords-Combat Robot; Wireless camer; Terror attack; Radio Operated; Self-Powered; Intruders

I. Introduction

The global focus on terrorism and security may have geared up following the 9/11 attacks in the USA. The risk of terrorist attack can perhaps never be eliminated, but sensible steps can be taken to reduce the risk. The word "Robot" was first used in a 1921 play titled R.U.R. Rossum's Universal Robots, by Czechoslovakian writer Karel Capek . Robot is a Czech word meaning "worker."

Merriam-Webster defines robot [2] as "**a machine that looks like a human being and perform various complex acts; a device that automatically performs complicated, often repetitive tasks; a mechanism guided by automatic controls**." ISO describes a robot as "**an automatically controlled reprogrammable, multipurpose manipulator programmable in three or more axes, which may be either fixed in place or mobile for use in industrial automation applications**".

Yet, all these definitions do give us a rough idea about what comprises a robot, which needs to sense the outside world and act accordingly. There are motors, pulleys, gears, gearbox, levers, chains, and many more mechanical systems, enabling locomotion. There are sound, light, magnetic field and other sensors that help the robot to collect information about its environment. There are Processors powered by powerful software that help the robot make sense environmental data captured and tell it what to do next and also microphones, speakers, displays, etc that help the robot interact with humans.

The main objectives of using robot are

A. Where man dares not venture

Robots have traditionally been put to use in environments that are too hazardous for man.

B. To rescue, pronto!

Robots also work under precarious conditions, for search and rescue after disasters. A host of robots built by the University of South Florida's Centre for robot assisted search and rescue were in action at the world trade centre site within hours after the disaster to delve into the rubble and rescue survivors. Similarly, robots are also put to work in underground mines. A lot of research today is focused on improving rescue functions of robots.

C. We even make them go to war

The faithful robots do not hesitate to tread even the dreaded terrain of battlefields [3]. Their use in Afghanistan and Iraq wars make us wonder if robots have indeed become intelligent! Battle robots of various shapes and sizes were deployed to defuse landmines, search for criminals hiding in caves, search for bombs under cars and in building, for espionage and what not! These robots were controlled by humans.

We aim to develop a model which will be efficiently used to minimize terrorist causality. Being able to achieve reliable long distance communication is an important open area of research to robotics as well as other technology areas. As interest in robotics continues to grow, robots are increasingly being integrated into everyday life. The results of this integration are end-users possessing less and less technical knowledge of the technologies [4].

Currently, the primary mode for robot communication uses RF. RF is an obvious choice for communication since it allows more information to be transferred at high speed and over long distance. This paper explores the use of readymade RF networks for communication and device control. This eliminates the need of a new infrastructure and detailed technical research.

II. Hardware implementation

The block diagram of the hardware implementation of the entire system is as shown in the Figure 1. This robot is radio operated, self-powered and has all the controls like a normal car. A pair of laser gun has been installed on it, so that it can fire on enemy remotely when required. Wireless camera will send real time video and audio signals, which could be seen on a remote monitor, and action can be taken accordingly.

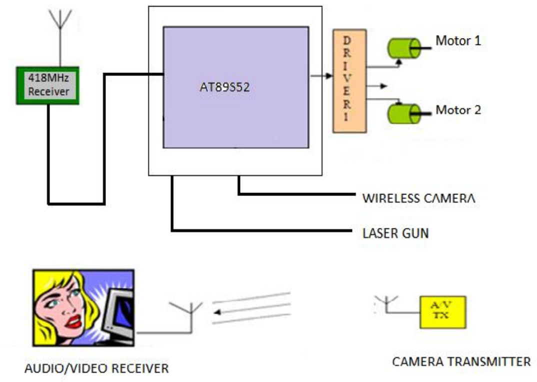


Figure 1: Block Diagram of Intelligent Combat Robot

Heart of our robot is Atmel's AT89S52 [5]. Microcontroller acts as master controller decodes all the commands received from the transmitter and give commands to slave microcontroller. It also acts as Slave microcontroller which is responsible for executing all the commands received from the master and also generating PWM pulses for the speed control. Based on the input codes master will give command to slave microcontroller and robot will behave as follows.

- moves in forward direction
- moves in reverse direction,
- speed controls in both the direction
- it can even turn left or right while moving forward or in reverse direction.
- Instant reverse or forward running without stopping.

A. *Transmitting unit*

Here a variable frequency oscillator 1 is used for modulating the frequency i.e. to be transmitted and has its output to a high frequency oscillator 2 for generating a carrier wave. The carrier wave is then radiated into space by the antenna. The transmitter module is shown in Figure 2.

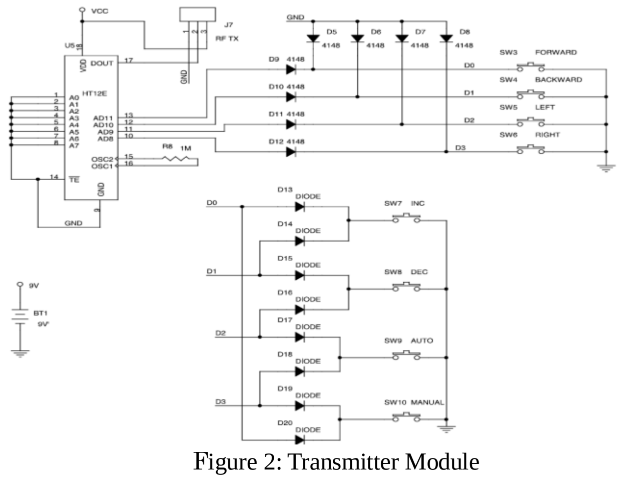


Figure 2: Transmitter Module

B. *Receiving Unit*

The receiving antenna is connected to a tuned wave detecting circuit for detecting the waves transmitted by transmitter antenna.The output of the tuned wave detecting circuit is connected to amplifier which in turn has its output connected to the input of the high pass frequency as well as the filter to a low pass frequency filter.

The outputs of amplifiers are connected to separate motors and other side of motors are connected to voltage potential .The high pass frequency filter extracts the higher frequency components of the output signals from the amplifier and the low pass frequency filter extracts the lower frequency components of the output signal from the amplifier .The receiver module is shown in Figure 3.

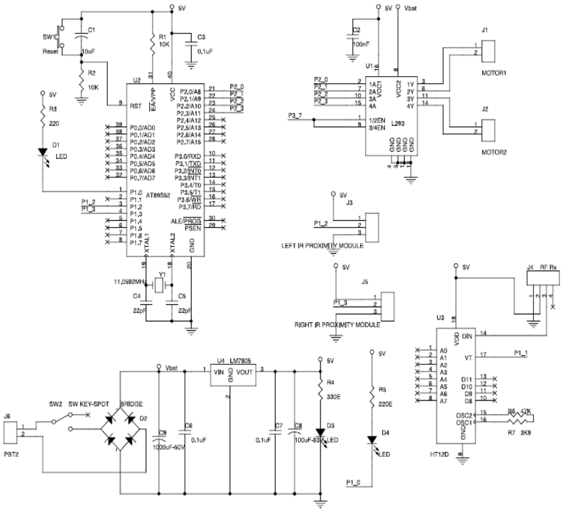


Figure 3: Receiver Module

III. Components or Subsystems Description

A. *Microcontroller circuit (AT89S52)*

It is the heart of the system which controls all the activities of transmitting and receiving. The IC used is AT89S52. The AT89S52 Microcontroller [6] is an 8-bit microcontroller with 8K Bytes of In-System Programming Flash Memory. The device is manufactured using Atmel's high-density nonvolatile memory technology and is compatible with the industry standard 80C51 instruction set and pin out. The on-chip Flash allows the program memory to be reprogrammed in-system or by a conventional nonvolatile memory programmer.

By combining a versatile 8-bit CPU with in-system programmable Flash on a monolithic chip, the Atmel AT89S52 is a powerful microcontroller which provides a highly-flexible and cost-effective solution to many embedded control applications. The Idle Mode stops the CPU while allowing the RAM, timer/counters, serial port, and interrupt system to continue functioning. The Power-down mode saves the RAM contents but freezes the oscillator, disabling all other chip functions until the next interrupt or hardware reset.

B. *Power supply circuit*

The main building block of any electronic system is the power supply to provide required power for their operation and is as shown in the Figure 4. For the microcontroller, keyboard, LCD, RTC, GSM, +5V are required & for driving buzzer +12V is required. The power supply [7] provides regulated output of +5V & non-regulated output of +12V. The three terminals IC7805 meets the requirement of +5V regulated. The secondary voltage from the main transformer is rectified by electronic rectifier & filtered by capacitor. This unregulated DC voltage is supplied to the input pin of regulator IC. The IC used are fixed regulator with internal short circuit current limiting & thermal shutdown capability.

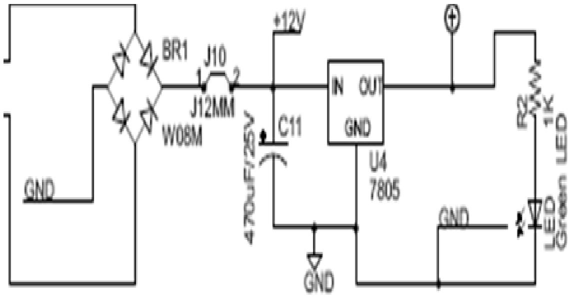


Figure 4: Power Supply Module

C. Decoder HT-12D

The decoders [8] are a series of CMOS LSIs for remote control system applications. They are paired with Holtek 2^{12} series of encoders. For proper operation, a pair of encoder/decoder with the same number of addresses and data format should be chosen.

The decoders receive serial addresses and data from a programmed 2^{12} series of encoders that are transmitted by a carrier using an RF or an IR transmission medium. They compare the serial input data three times continuously with their local addresses. If no error or unmatched codes are found, the input data codes are decoded and then transferred to the output pins.

The VT pin also goes high to indicate a valid transmission. The 2^{12} series of decoders are capable of decoding information's that consist of N bits of address and 12_N bits of data. Of this series, the HT12D is arranged to provide 8 address bits and 4 data bits, and HT12F is used to decode 12 bits of address information

D. Encoder HT-12E

The 2^{12} encoders are a series of CMOS LSIs for remote control system applications. They are capable of encoding information which consists of N address bits and 12_N data bits. Each address/data input can be set to one of the two logic states. The programmed addresses/data are transmitted together with the header bits via an RF or an infrared transmission medium upon receipt of a trigger signal. The capability to select a TE trigger on the HT12E [8] further enhances the application flexibility of the 2^{12} series of encoders.

E. DC Motors

For the movement of our robot, we are using DC motors [9]. It is operated by 12V DC power supply. In any electric motor, operation is based on simple electromagnetism. A current carrying conductor generates a magnetic field; when this is then placed in an external magnetic field, it will experience a force proportional to the current in the conductor, and to the strength of the external magnetic field.

F. Motor Driver L293D

The Device is a monolithic integrated high voltage, high current four channel driver designed to accept standard DTL or TTL logic levels and drive inductive loads and switching power transistors. To simplify use as two bridges each pair of channels is equipped with an enable input.

A separate supply input is provided for the logic, allowing operation at a lower voltage and internal clamp diodes are included. This device is suitable for use in switching applications at frequencies up to 5 kHz. The L293D is assembled in a 16 lead plastic package which has 4 center pins connected together and used for heat sinking. The chip is designed to control 2 DC motors. There are 2 Input and 2 output pins for each motor. The behavior of motor for various input is shown in Table 1.

TABLE 1. BEHAVIOR OF MOTORS

Operation	**A**	**B**
Stop	Low	Low
Clockwise	Low	High
Anti Clockwise	High	Low
Stop	High	High

G. Transmitter for Laser Gun

The transmitter is constituted by AT90S2323 microcontroller and TLP434 RF transmitter module at 418MHz. Transmitter is designed for more battery economy and safe transmission of the data.

Block diagram for the transmitter of laser gun is as shown in the Figure 5.

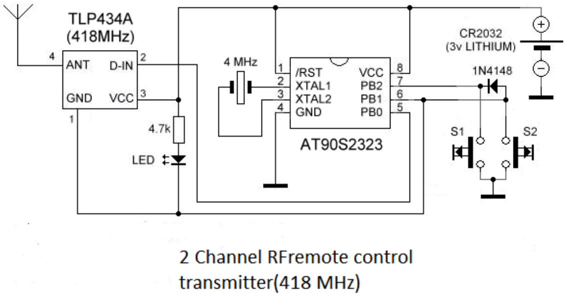


Figure 5: Block Diagram of Transmitter Laser Gun

Here TLP434A is an Ultra-small Trasmitter of range 418MHz with ASK modulation scheme with operating voltage range from 2-12 dc voltage. This IC is usually chained with the encoder IC for example HT12-E. This transmitter is connected to the AT90S2323 10MHz with 2k flash microcontroller. This constitutes a transmitter section of laser gun.

H. Receiver for Laser Gun

The receiver constituted by RF receiver module RLP434A at 418MHz, the microcontroller AT90S2313 [10] and the two relays which can handle any electric (or electronic) device up to 10 Amps (the contacts of my relays are 10Amp at 250Volts). The RLP434A is an RF receiver module with receipt frequency at 418MHz with ASK modulation. There are 2 outputs from this module, the digital, with levels from 0v to VCC (5 volts in our case) and the analog output. Analog output is not used. The transmitter sends 4 bytes with 2400bps 4 times and the receiver RLP-434A collects them and moves them to AT90S2313 to RxD pin, PD0. Block diagram for receiver of the laser gun is depicted in Figure 6.

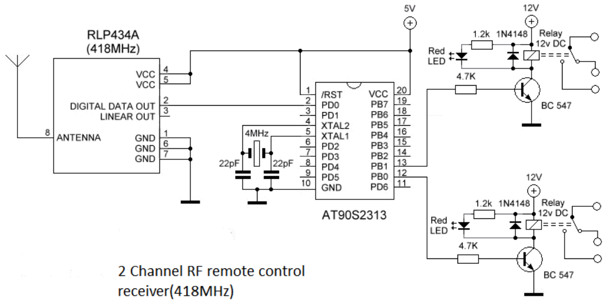


Figure 6: Block Diagram of Receiver Laser Gun

Here RL434A is SAW (surface acoustic wave) based receiver compatible of 418MHz of frequency with operating range from 3.3-6 dc voltage and also it employs ASK modulation. Again as a Tx, here also we have 8-bit microcontroller AT90S2313 with 2k flash memory with 11 pin DIP . All these components will constitute a receiver of laser gun.

I. RF Communication

Radio frequency (RF) is a rate of oscillation in the range of about 3 kHz to 300 GHz, which corresponds to the frequency of radio waves, and the alternating currents which carry radio signals. RF usually refers to electrical rather than mechanical oscillations. The energy in an RF current can radiate off a conductor into space as electromagnetic waves (radio waves); this is the basis of radio technology.

J. JMK AV Receiver with Wireless Camera

It is mini wireless monitoring video camera and wireless receiver set for home and small business surveillance and is used here for demonstration purpose. Simply install the wireless camera in the room where we want to monitor and set the wireless receiver in the next room (up to 15 meters away) and hook it up to a TV or DVR to watch the action or record the footage for the security records.

Here we are placing this wireless camera in the combat robot. Depiction of AV Receiver wireless camera is as shown in Figure 7.

K. TV Capture card

A TV capture card is a computer component that allows television signals to be received by a computer. It is a kind of television tuner. Most TV tuners also function as video capture cards, allowing them to record television programs onto a hard disk. Digital TV tuner card is as shown in the Figure 8.

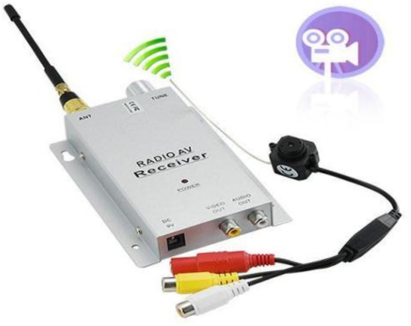


Figure 7: AV Receiver and Wireless Camera

The card contains a tuner and an analog-to-digital converter along with demodulation and interface logic.

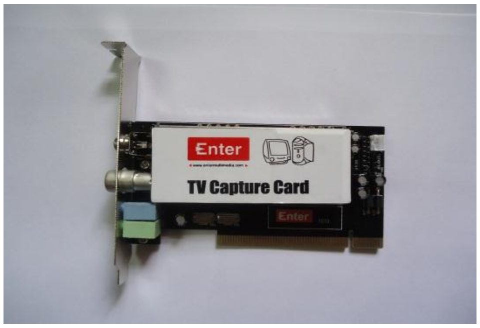


Figure 8: ATI digital TV capture card

L. Remote Controller Decoder SC2272-T4

It can be used for wireless remote control receivers and its features are

- Operating voltage: DC 4~6V.
- Up to 12 tri-state code address pins.
- Up to 6 data pins.
- Toggle control mode.

IV. SOFTWARE IMPLEMENTATION

For the software implementation, we deploy two software packages. First one is the Keil µVision 3.0. second one is the Flash magic simulator.The Keil µVision Debugger accurately simulates on-chip peripherals (I²C, CAN, UART, SPI, Interrupts, I/O Ports, A/D Converter, D/A Converter, and PWM Modules) of 89S52 device.

Simulation helps to understand hardware configurations and avoids time wasted on setup problems. With simulation, we can write and test applications before target hardware is available. The system program written in embedded C [11] using KEIL IDE software will be stored in Microcontroller.

Keil development tools for the Microcontroller Architecture support every level of software developer from the professional applications engineer to the student for learning about embedded software development. The industry-standard Keil C Compilers, Macro Assemblers, Debuggers, Real-time Kernels, Single-board Computers, and Emulators support all 89S52 derivatives. The Keil Development Tools are designed to solve the complex problems facing embedded software developers.

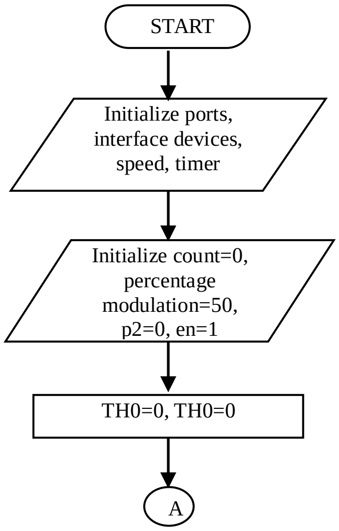
Flash magic is used to dump the code to microcontroller from PC. Flash Magic is a free, powerful, feature-rich Windows application that allows easy programming of Philips FLASH Microcontrollers. Build custom applications for Philips Microcontrollers on the Flash Magic platform! Use it to create custom end-user firmware programming applications, or generate an in-house production line programming tool.

The Flash Memory In-System Programmer is a tool that runs under Windows 95/98/NT4/2K. It allows in-circuit programming of FLASH memories via a serial RS232 link. Computer side software called Flash Magic is executed that accepts the Intel HEX format file generated from compiler Keil to be sent to target microcontroller. It detects the hardware connected to the serial port.

A. Flow charts

The flowcharts depicting the Robot Movement and its Delay are shown in Figure9, 10 and 11.

1) Robot Movement


START
Initialize ports, interface devices, speed, timer
Initialize count=0, percentage modulation=50, p2=0, en=1
TH0=0, TH0=0
A

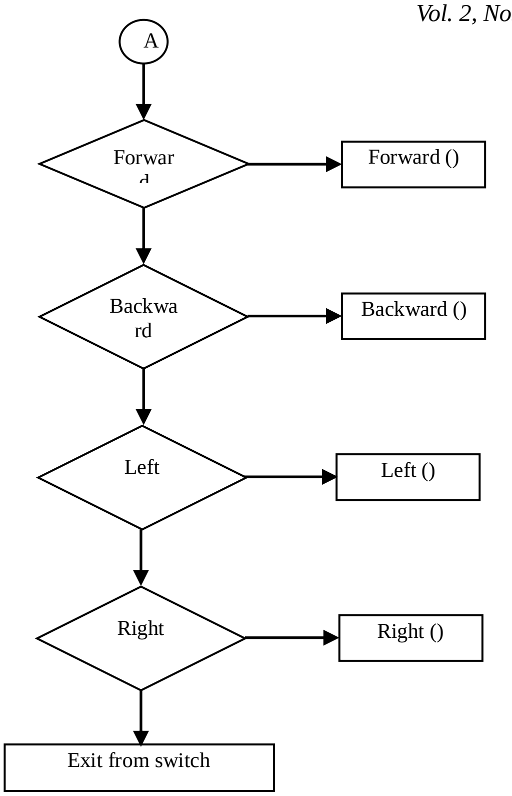


Figure 9: Flowchart For Robot Movement

2) Robot for Particular Movement

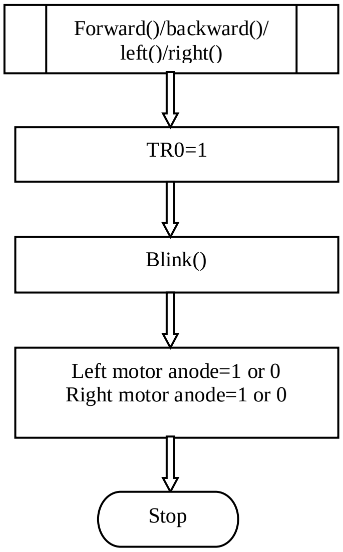


Figure 10: Flowchart For Particular Movement

3) Delay Flowchart

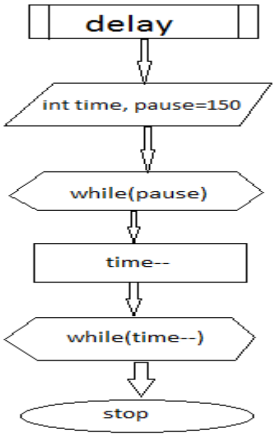


Figure 11: Flowchart For Robot Movement

V. Results and Discussion

Remote controllers are designed to direct the orientation of robot and to operate the laser gun. Robot keeps on moving in two modes i.e., Manual mode and self-mode. It’s brought under user’s control in the case of manual mode. In self-mode, robot starts moving over surface and takes action according to the scenario. To detect the obstacles, we have deployed Infrared sensors (left sensor and right sensor) in the front portion of the module. While moving on the surface, if the left sensor is detected, robot takes back the position for a moment and moves right. If the right sensor is detected, robot gets back and moves left. The front view and top view of designed combat robots are shown in the figures 12 &13.

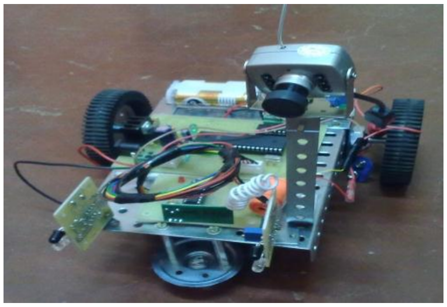


Figure 12: Front view of designed combat robot

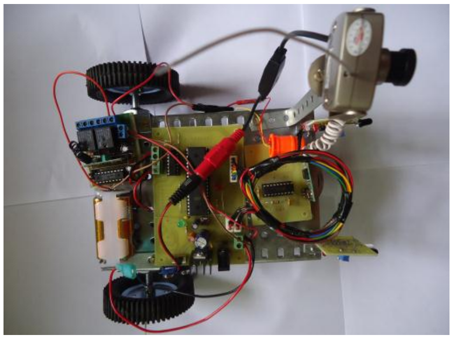


Figure 13: Top view of designed combat robot

VI. Applications

- Can be adequately implemented in national defense through military-industrial partnership. It is shown in the figure 14.

Figure 14: Top view of combat robot

- Can be vastly applied in Resorts, borders of noted buildings. It is shown in the figure 15.

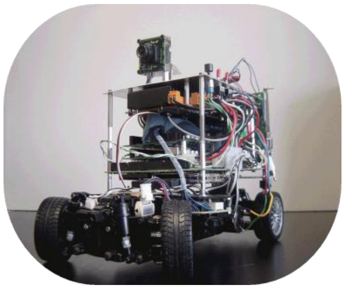


Figure 15: Top view of combat robot

- Installation of combat robots in the stadiums, sacred places, government and non government organizations assures top security.

VII. Conclusion

As we all know, these days India is sick off massive terror attacks, bomb explosions at plush resorts. To avoid such disasters TECHNOLOGICAL power must exceed HUMAN power. Human life and time are priceless.

It’s our onus to take an initiative to design a model of an apt robot that meets combatant needs. So to avoid terror attacks, to ensure more security at the border and high density areas it’s wise to maintain a world class military technology in accordance with combatant needs.

Even every nation needs its own defense system for their integrity and security. In such a way construction of these robots will carry nation’s name, fame globally.

References

[1] Pete Miles & Tom Carroll, Build Your Own Combat Robot, (2002).

[2] K.S.Fu , R.C.Gonzalez , C.S.G..Lee, Tutorials Robotics.

[3] Asaro,P. How just could a robot war be?, Frontiers in Artificial Intelligence and Applications, 75, 50-64.

[4] S. Y. Harmon & D. W. Gage, “Current Technical Research Issues of Autonomous Robots Employed In Combat”, 17th Annual Electronics and Aerospace Conference.

[5] www.Atmel.com

[6] Atmel data sheets http://www.keil.com/dd/docs/datashts/atmel/at89s52_ds.pdf

[7] Robert L.Boylestad and Louis Nashelsky, "Electronic Devices and Circuit Theory", 8th Edition, 2006

[8] Decoder HT-12D, Encoder HT-12E http://robokits.co.in/shop/index.php?main_page=product_info&cPath=14_15&products_id=76>

[9] A. Khamis, M. Pérez Vernet, K. Schilling, "A Remote Experiment On Motor Control Of Mobile Robots", 10thMediterranean Conference on Control and Automation – MED2002.

[10] Receiver for Laser Gun www.alldatasheet.com/datasheet-pdf/pdf/169605/.../RLP434A.htm

AUTHORS PROFILE

Dr.S.Bhargavi is presently working as a Professor in the department of Electronics and Communication engineering, SJCIT, Chikballapur, Karnataka, India. She is having 12 years of teaching experience. Her areas of interest are Robotics, Embedded Systems, Low Power VLSI, Wireless communication, ASIC and Cryptography.

S.Manjunath is presently working as a Lecturer in the department of Electronics and Communication engineering, SJCIT, Chikballapur, Karnataka, India. He is having 3 years of Industrial experience and 1 year of teaching experience. His areas of interest are Robotics, Embedded Systems and Wireless communication.

Analysis of Guess and Determined Attack on Non Linear Modified SNOW 2.0 Using One LFSR

Madiha Waris [#1], Malik Sikandar Hayat Khiyal[#2] , Aihab Khan [#3]
Department of Software Engineering Fatima Jinnah Women University, Old Presidency, The Mall, Rawalpindi, Pakistan

***Abstract*—stream ciphers encrypt the data bit by bit. In this research a new model of stream cipher SNOW 2.0 has been proposed i.e. Non linear modified SNOW 2.0 using one Linear Feedback Shift Register (LFSR) with the embedding of non linear function in the model. The analysis of Guess and Determined (GD) attack has been done to check its security with respect to previous versions. The proposed model contains one Linear Feedback Shift Register (LFSR) along with the non linear function which increases the strength of the stream cipher, to make the static nature of modified SNOW 2.0 dynamic. The Experimental analysis show that such a mechanism has been built which provides more security than the previous version of modified SNOW 2.0 in which non linearity was either not introduced or it was introduced using two Linear Feedback Shift Registers (LFSRs). It is concluded that this version is more powerful with respect to the security of plain text against Guess and Determined attack (GD) as compared to the previous versions.**

Keywords-component; Linear Feedback Shift Register; Guess and Determined Attack; Finite State Machine .

I. INTRODUCTION

A stream cipher is a cryptographic way to achieve the confidentiality on communication channel [1]. They are broadly used for the provision of security for communication and in network streams.

Linear Feedback Shift Register (LFSR) is the frequently used device as key stream generator. It is mostly used in many key stream generators due to its simplicity. The SNOW family is a typical example of word oriented stream cipher, based on Linear Feedback Shift Register [2].

The initial version of SNOW was presented to NESSIE project in 2000, which wasn't accepted due to the weaknesses present in it against the Guess and Determined (GD) attack. Then the new version SNOW 1.0 was developed by Thomas Johansson and Patrik Ekdahl and it was also discarded due to Guess and Determined (GD) attack. Then these authors presented a new version SNOW 2.0 which was the modified version of SNOW 1.0 with the enhancement of features for secure communication. It was claimed to solve the weaknesses and improvement of the performance. Later on a Modified version of SNOW 2.0 was built by Hadi Ahmadi which was static in nature and it was also less secure against the Guess and Determined (GD) attack [3].

In this paper, the static feedback based modified SNOW 2.0 has been converted into dynamic feedback based modified SNOW 2.0 by using one Linear Feedback Shift Register (LFSR) and with the addition of non linear function which enhances the security performance of the model by using irregular clocking. This will increase the complexity for the attacker to guess the input. The proposed model of Non linear SNOW 2.0 with one Linear Feedback Shift Register (LFSR) is checked for security against Guess and Determined (GD) attack and by experimental analysis it is observed that it is more secure.

The paper is organized in such a way that: section 2 discusses related work to proposed schemes, section 3 demonstrates the proposed framework, section 4 discusses proposed technique along with algorithm ,section 5 discusses analysis of proposed technique, section 6 describes the experimental results and at the end the conclusion is given in section 7.

II. RELATED WORK

Ahmadi and Salehani et al. [3] proposed a Modified Version of SNOW 2.0. He gave a criterion of modifying an LFSR-based stream cipher against Guess and Determined (GD) attacks. In this model a stream of pseudorandom digits in a synchronous stream cipher was independent of the plaintext and cipher text messages, and then combined with the plaintext for encryption or with the cipher text for decryption. IrfanUllah and Naz et al. [4] proposed a model of SNOW 2.0 and claimed that it contained a series of patterns of bits that were traceable. Two versions SNOW 1.0 and SNOW 2.0 were examined and concluded that if the plaintext that has to be encrypted is of small amount, modified Version of SNOW 2.0 should be used and if large data set has to be encrypted original snow 2.0 should be recommended. Khan et al. [2] proposed the model of dynamic feedback based modified SNOW 2.0 and compared the dynamic nature of modified SNOW 2.0 with the previous version of modified SNOW 2.0 that was static in nature. Also compared them on the basis of Guess and Determined (GD) attack and concluded that Dynamic Feedback based Modified SNOW 2.0 is more secure than static feedback based modified SNOW 2.0 for the encryption of plain text. Masood et al. [5] proposed the model for analysis of non linear Snow 2.0 for improved security and did it by embedding a non linear function using two Linear Feedback Shift Registers (LFSRs) in dynamic feedback based modified SNOW 2.0.

In this technique the linear behavior of static feedback based modified SNOW 2.0 has been converted to non linear behavior by introducing non linear function based on irregular clocking. For this the feedback change accepts values at dynamic tap positions rather than static so its structure is considered as dynamic and non linear and it was found that this dynamic feedback mechanism for Linear Feedback Shift Register (LFSR) and nonlinear behavior was an effective method to improve the security of SNOW 2.0 and hence it resulted in increased complexity for attacker to guess the input.

III. Proposed Framework

The proposed model comprises dynamic feedback based modified SNOW 2.0 with one Linear Feedback Shift Register (LFSR) along with the non linear function. For security provision in stream ciphers dynamic feedback mechanism is the best way as it changes a deterministic linear recurrence of some registers into probabilistic recurrence [2]. Due to this characteristic the stream cipher having dynamic feedback control mechanism remain protected against various attacks.

The attacker has to guess the some inputs to the non linear function for which he has to obtain a linear recurrence of the key stream. The irregular modification makes it impossible for the attacker to find the linear recurrence of the key stream obtained by some registers. In this way the security enhances.

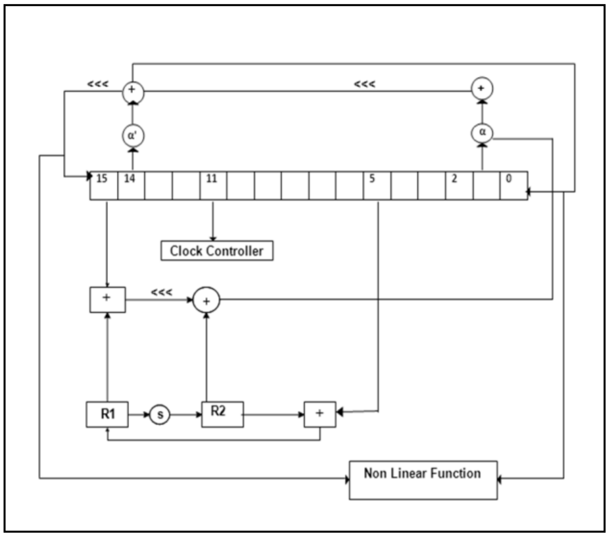


Figure 1. Proposed Model for Non linear Modified SNOW 2.0 with one LFSR

The Proposed model as shown in “Fig. 1”is constructed in such a way that it consists of one Linear Feedback Shift Register (LFSR), Finite State Machine (FSM) and a Non linear function having a dynamic feedback controller and two internal registers M1 and M2.

A. *Symbols Specification of the Proposed Model*

R1= 32 bit Register	$\oplus$ = bit wise XOR
R2= 32 bit Register	S= S-box
<<<= cyclic shift 7 steps left	$\boxplus$ =addition modulo 2^{32}

B. *Non linear function for proposed non linear SNOW 2.0.*

The non-linear function of proposed model as shown in “Fig. 2” is fed the values of four of the tap positions of Linear Feedback Shift Register (LFSR) that are dynamically taken into R1, R2, L1, and L2 which are internal memories and outputs 64 bits of key-stream for every cycle.

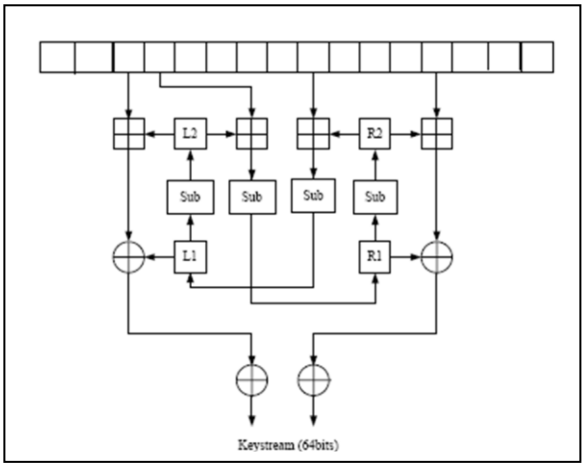


Figure 2. Nonlinear function model description using one LFSR.

IV. Proposed Technique

A. *Working Scenerio of the Proposed Model*

The working of the proposed model is described in detail in this section. The whole process continues in such a way that:

1. At first the key initialization is done in order to initialize LFSR. The registers R1 and R2 of the Finite State Machine (FSM) are set to zero.
2. In the next step, the cipher clocks 32 times without producing any output and the FSM output is fed back into the Linear Feedback Shift Register (LFSR).
3. The steps for the proper working of cipher are as follows:
   - $\text{tempf}_t = S_{t+15} \boxplus R1_t$
   - Left shift circular $S_{t+15} \boxplus R1_t$, and then Xor it with $R_{2.}$
   - $\text{tempf}_t = (S_{t+15} \boxplus R1_t) <<< 7$
   - $f_t = \text{tempf}_t \oplus R2_t$; when t≥0
4. The non linear function is implemented in such a way that at first initial vector (IV) values are initialized and a key is loaded. The dynamic tap positions are taken and the whole process continues as follows:
   - $f_{t1} = \text{temp} \oplus R3_t$; when t≥0
   - keystream 1= $f_{t1} \oplus$ dynamic no.
   - $f_{t2} = \text{temp1} \oplus L_2$
   - keystream 2= $f_{t2} \oplus$ dynamic no.
5. The final keystream is determined as follows:
   Final keystream = keystream 1 $\oplus$ keystream 2
6. The next state of Finite State Machine (FSM) registers is determined as follows:

- $R1_{t+1} = S_{t+5} \boxplus R2_t$ and
- $R2_{t+1} = S(R1_t)$ $t \geq 0$

7. The next state of Linear Feedback Shift Register (LFSR) is determined as follows:
$S_{16} = ((\alpha^{-1} S_{t+11}) <<<,7) \oplus S_{t+2} \oplus ((\alpha S_t) <<<, 7)$

B. Graphical Representation of the Proposed Model

The graphical model of the Non linear modified SNOW 2.0 using one Linear Feedback Shift Register (LFSR) is shown in "Fig. 3". It depicts the flow of the whole process. At first the dynamic number generator will generate dynamic numbers which will be fed to the shift registers.

The values form the shift registers will be then given to the clock-controller and the values of shift registers will be updated. The updated values will then be fed to the non linear function which will generate key stream.

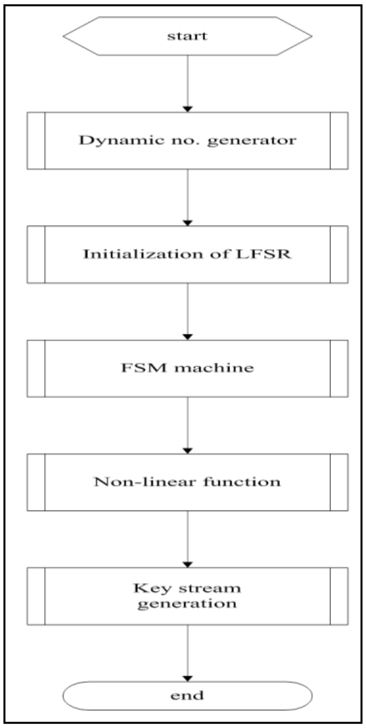


Figure 3. Graphical Model for Non linear Modified SNOW 2.0 with one LFSR

C. Algorithm of the Proposed Technique

1) Algorithm: Generation of Dynamic numbers

Input: Time

Output: srand

Read time.

srand= srand (time(NULL))

srand= return (rand()%16+1)

output srand.

2) Algorithm: Shift Registers

Input: IV values

Output: Keystream

Read initial values

ptemp = snow r1+*(snow_ptr+15)

ptemp1= ptemp<<7

snow_outfrom_fsm=ptemp^snow_r2

Output keystream.

3) Algorithm: Clocking

Input: FSM output

Output: no output

Read snow_outfrom_fsm

Update Interanls

No output.

4) Algorithm: Non linear Function

Input: Four numbers generated dynamically

Output: keystream

Read dynamic numbers

Add them to internal memories

Xor the resultant values with dynamic numbers

output keystream.

V. Analysis of Proposed Technique

The performance of proposed technique can be analyzed by the help of attack. Guess and Determined (GD) attack has been done on the proposed technique in order to measure its security position.

Guess and determined (GD) attacks have been affecting some stream ciphers. In this attack, the attacker judges the correlation between the cipher's building blocks [6].

In this attack the attacker at first guesses the initial states of the cipher, which is known as a basis. Then he finds the running key sequence. If his sequence matches the original key sequence it means that he has guessed the original key stream and if it does not match then he tries again to guess with new initial values and key until he finds the original key stream.

This process is based on the fact that, the attack complexity will be less if the basis size is small [7]. Due to presence of the nonlinear function, it is not possible for the attacker to guess the correct values.

The Guess and Determined (GD) attack occurs in the sequence as follows [8]:

a) The attacker guesses initial values for the Finite State Machine (FSM)

b) The attacker guesses the values for registers *c)* The values obtained from registers are used by the attacker to guess the Linear Feedback Shift Register (LFSR) state.

d) The attacker generates a key stream and tests the values of Linear Feedback Shift Register (LFSR) and Finite State Machine (FSM) state and then he compares the produced key stream with the original key stream.

In the proposed system the attack is applied in this way that at first the initial values and secret key is guessed. Then the key is converted into binary form. This binary format is then converted into 32 bit format. Then key stream is determined with the help of this secret key initialization. At the end this key stream is compared with the original key stream.

VI. Experiment results and Analysis

The Guess and Determined (GD) attack has been applied on the proposed model and on the previous versions also to check the security standard of the proposed version. The comparison algorithm has been applied to compare the key streams generated with the guessed key streams. It works in such a way that the result is "1" if the key stream matches and it results "0" if the key stream is not matched.

Following are the two phases in which the results have been examined.

A. *Analysis of Phase I*

In order to determine the original key stream on basis of guess, 30 experiments have been performed in Phase I. 10 experiments are performed on each of Non linear Modified SNOW 2.0 using one Linear Feedback Shift Register (LFSR), Dynamic feedback based modified snow 2.0 using non-linear function, dynamic feedback based modified snow 2.0 and static feedback based modified snow 2.0. 1 attack is applied on each experiment and each attack has 10 guesses.

One attack generates 50 guess key streams; mean 10 attacks generate 500 key streams. Hence, 30 experiments have 1500 guesses in total.

Experimental analysis show that the proposed model show more resistance to Guess and Determined (GD) attack as compared to the previous versions. Results of the Phase I are shown in "Table I".

TABLE I. Evaluation of Phase I

Expe rime nts	Non linear SNOW 2.0 using one LFSR	Non linear SNOW 2.0 using 2 LFSRs	Dynamic feedback based modified SNOW 2.0	Static feedback based modified SNOW 2.0
	Similarities	**Similarities**	**Similarities**	**Similarities**
1	375	381	397	406
2	351	345	400	410
3	405	409	439	456
4	357	357	434	466
5	382	380	409	471
6	360	364	392	495
7	404	407	420	452
8	372	390	376	415
9	387	396	456	442
10	363	363	456	404
Total	3756	3792	4179	4417

According to "Table I" the Non linear SNOW 2.0 using one Linear Feedback Shift Register (LFSR) is more secure as compared to the other defined versions of SNOW 2.0.

The graphical representation of Phase I is shown in "Figure 4".

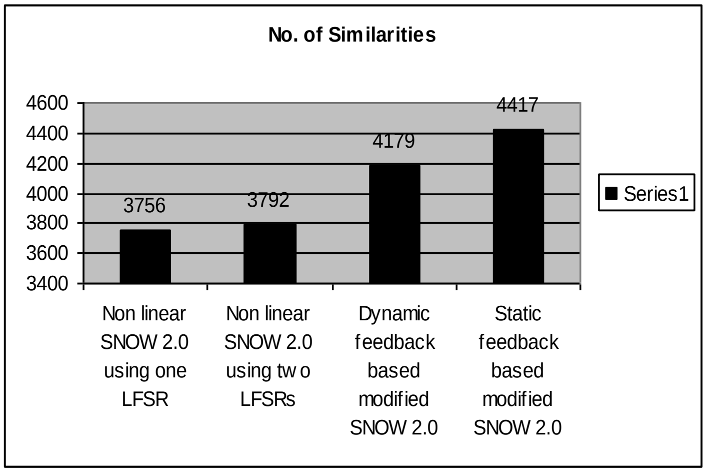


Figure 4. Graphical Representation of Phase I

B. *Analysis of Phase II*

In Phase II, 30 experiments have been performed for the analysis of key streams, each experiment has 10 attacks and each attack has 50 key streams. Similar to Phase I, 10 experiments contains 500 guesses.

10 experiments are performed on each of Non linear Modified SNOW 2.0 using one Linear Feedback Shift Register (LFSR), dynamic feedback based modified snow 2.0 using non-linear function, dynamic feedback based modified snow 2.0 and static feedback based modified snow 2.0. The only difference from Phase I is that, in this phase 40 key streams have been generated against each guess. As the number of key streams increase, the number of comparisons also increases.

Experimental analysis show that the proposed model show more resistance to Guess and Determined (GD) attack in Phase II as compared to the previous versions. Results of the Phase II are shown in "Table II".

TABLE II. EVALUATION OF PHASE II

Experi ments	Non linear SNOW 2.0 using one LFSR	Non linear SNOW 2.0 using 2 LFSRs	Dynamic feedback based modified SNOW 2.0	Static feedback based modified SNOW 2.0
	Similarities	**Similarities**	**Similarities**	**Similarities**
1	818	840	863	851
2	759	816	854	825
3	780	728	803	830
4	807	827	84	938
5	751	822	867	907
6	762	810	859	943
7	814	830	864	903
8	783	798	874	901
9	733	736	789	934
10	750	806	865	934
Total	7757	8013	8512	8966

According to “Table II” the Non linear SNOW 2.0 using one LFSR is more secure as compared to the other defined versions of SNOW 2.0. The graphical representation of Phase I is shown in “Figure 4”.

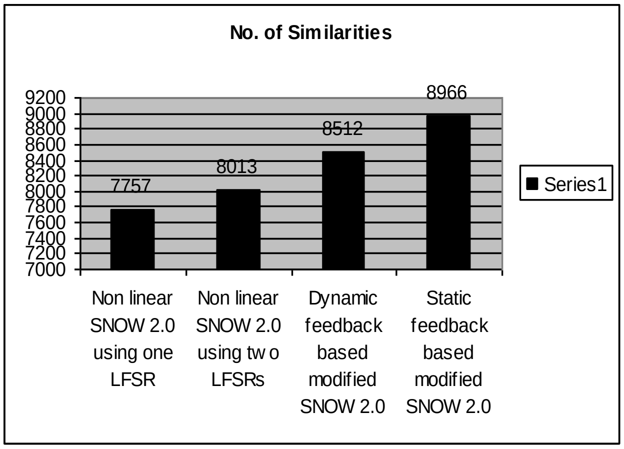


Figure 5. Graphical Representtion of Phase II

VII. CONCLUSION

In this paper the Guess and Determined (GD) attack has been applied on Non linear modified SNOW 2.0 using one Linear Feedback Shift Register (LFSR). The results concluded from Phase I and Phase II by comparison algorithm show that the proposed technique is more powerful with respect to security as compared to Non linear SNOW 2.0 using two Linear Feedback Shift Registers (LFSRs), Dynamic feedback based modified SNOW 2.0 and Static feedback based modified SNOW 2.0. The use of non linear function with one LFSR is a more reliable method with the reduction of complexity as compared to previous versions of SNOW 2.0.

REFRENCES

[1] P. Ekdahl, T. Johansson, *"A new version of the stream cipher SNOW"*, Proceedings of Selected Areas in Cryptography (SAC) 2002, Volume 2595 of LNCS, pages 47-61, Springer, 2002.

[2] Saira Khan, Aihab Khan, Malik Sikandar Hayat Khiyal, Tarranum Baz “Dynamic Feedback based Modified SNOW 2.0” IEEE International Conference on Emerging Technologies 2010 (ICET 2010), October 18-19, 2010, Islamabad, Pakistan.

[3] Ahmadi H., Esmaeili Salehani Y., "A Modified Version of SNOW2.0", International CSI Computer Conference, 2007

[4] Syed IrfanUllah, Tarannum Naz and Sikandar Hayat Khiyal “ Traceable Bit Streams in SNOW 2.0 using Guess-and-Determine Attack” World Applied Sciences Journal, Volume 11, No. 2, pp 190-195, 2010.

[5] Mina Masood, Aihab Khan, Malik Sikandar Hayat Khiyal, Ghoosia Arshad “ Analysis and Design of Non-Linear SNOW 2.0 for improved security” International Journal of Computer Technology and Engineering. (Submitted)

[6] Rohani, N.; Noferesti, Z.; Mohajeri, J.; Aref, M.R.; , "Guess and Determine Attack on Trivium Family," Embedded and Ubiquitous Computing (EUC), 2010 IEEE/IFIP 8th International Conference, pp.785-790, 11-13 Dec. 2010

[7] Ahmadi, H.; Eghlidos, T.; , "Heuristic guess-and-determine attacks on stream ciphers," Joual of IET- Information Security, Institute of Engineering and Technology, vol.3, Issue. no.2, pp.66-73, June 2009

[8] P. Hawkes and G. G. Rose. Guess-and-determine attacks on SNOW. In K. Nyberg and H. Heys, editors, *Selected Areas in Cryptography*- SAC 2002, volume 2595 of *Lecture Notes in Computer Science*, pages 37-46. Springer Verlag, 2002.

AUTHOR’S PROFILE

Madiha Waris is a graduate from Dept. of Software Engineering, Fatima Jinnah Women University, Pakistan.

Dr. M. Sikandar H. Khiyal born at Khushab, Pakistan. He is Chairperson Dept. Computer Sciences and Software Engineering in Fatima Jinnah Women University Pakistan. He served in Pakistan Atomic Energy Commission for 24 years and involved in different research and development program of the PAEC. He developed software of underground flow and advanced fluid dynamic techniques. He was also involved at teaching in Computer Training Center, PAEC and International Islamic University. His area of interest is Numerical Analysis of Algorithm, Theory of Automata and Theory of Computation. He has more than one hundred research publications published in National and International Journals and Conference proceedings. He has supervised more than one hundred and thirty research projects at graduate and postgraduate level. He is member of SIAM, ACM, Informing Science Institute, IACSIT. He is Associate editor of IJCTE, and International Journal of Reviews in Computing. He is reviewer of the journals, IJCSIT, JIISIT, IJCEE, JCIE and CEE of Elsevier.

Mr.Aihab Khan works in Dept. of Computer Sciences at Fatima Jinnah Women University, Pakistan. His research interests are in the field of Data Mining, Data Warehousing as well as Information security.

Identification Problem of Source Term of A Reaction Diffusion Equation

Bo Zhang
Linyi University at Feixian
Feixian, Shandong, P.R.China

Abstract—**This paper will give the numerical difference scheme with Dirichlet boundary condition， and prove stability and convergence of the difference scheme, final numerical experiment results also confirm effectiveness of the algorithm.**

Keywords- Fractional derivative; Numerical difference scheme; The gradient regularization method.

I. INTRODUCTION

Source term inversion in groundwater pollution is a class of inverse problems[7,8], and is also important field of inverse problem research. Scholars have done a large amount of work and obtained many results. Reference [2] presented a new gradient regularization algorithm to solve an inverse problem of determining source terms in one-dimension solute transport with final observations, and reference [3] proposed a implicit method to solve a class of space-time fractional order diffusion equations with variable coefficient.

However, when fractional derivative replaces second derivative in diffusion equations, there is anomalous diffusion phenomenon. In this paper, we give the numerical difference scheme in source term identification with Dirichlet boundary condition, and we prove the stability and convergence of the difference scheme, also verify the practicality and effectiveness of the algorithm through numerical experiment.

In this paper, we use difference scheme to solve the forward problem, and when solving the inverse problem, we use the gradient regularization method based on Tikhonov regularization strategy. Here, the additional information for source term identification is set as the final observations, and suppose that the source term function are only concerned with the space variable and has nothing to do with the time variable.

In fact, the solute transport model can be described by the following equation[1]

$$\frac{\partial y}{\partial t} = u(x)\frac{\partial^2 y(x,t)}{\partial x^2} - v(x)\frac{\partial y(x,t)}{\partial x} - c(x)y(x,t) + b(x,t), \quad x \in [0,L], t \in [0,T], \tag{1}$$

by introducing fractional derivative and adding initial boundary conditions[5], Eqs. (1) will be modified as the following problem

$$\frac{\partial^\alpha y}{\partial t^\alpha} = u(x)\frac{\partial^\beta y(x,t)}{\partial x^\beta} - v(x)\frac{\partial y(x,t)}{\partial x} - c(x)y(x,t) + b(x,t), \qquad x \in \Omega, t > 0, \tag{2}$$

$$Dy(x,t) = g(x,t), \qquad x \in \Omega, t > 0, \tag{3}$$

$$Ey(x,0) = f(x), \qquad x \in \Omega, \tag{4}$$

where $0 < \alpha < 1, 1 < \beta < 2$, D represents matrix operator with boundary condition, E represents matrix operator with initial condition, $y(x,t)$ and $b(x)$ represent undetermined source term function and undetermined vector function respectively.

Inverse problem of this problem is to determine unknown vector function $b(x)$ by Eqs. (2)-(4) and the additional condition below

$$y(x,t)|_{x=T} = \psi(x). \tag{5}$$

II. NUMERICAL DIFFERENCE SCHEME

Considering the following space-time reaction diffusion equations

$$\frac{\partial^\alpha y(x,t)}{\partial t^\alpha} = u(x)\frac{\partial^\beta y(x,t)}{\partial x^\beta} - v(x)\frac{\partial y(x,t)}{\partial x} - c(x)y(x,t) + b(x,t), \tag{6}$$

$$y(x,0) = f(x), \tag{7}$$

$$y(0,t) = g_1(t), y(L,t) = g_2(t). \tag{8}$$

Where $u(x) > 0, v(x) > 0, c(x) \ge 0$ are continuous functions on $[0,L]$, $g_1(t) > 0, g_2(t) > 0$ are continuous functions on $[0,T]$, $b(x,t)$ is continuous on $[0,L] \times [0,T], \alpha \in (0,1), \beta \in (1,2), \frac{\partial^\alpha y(x,t)}{\partial t^\alpha}, \frac{\partial^\beta y(x,t)}{\partial x^\beta}$ are

Caputo fractional derivative and Riemann-Liouville fractional derivative respectively[5]

$$\frac{\partial^{\alpha} y(x,t)}{\partial t^{\alpha}} = \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-\eta)^{-\alpha} \frac{\partial y(x,\eta)}{\partial \eta} d\eta,$$
(9)

$$\frac{\partial^{\beta} y(x,t)}{\partial x^{\beta}} = \frac{1}{\Gamma(2-\beta)} \frac{\partial^2}{\partial x^2} \int_0^t \frac{y(\xi,t)}{(x-\xi)^{\beta-1}} d\xi.$$
(10)

Suppose that Eqs. (6)-(8) has a unique and smooth enough solution. $\tau = T/n$ is time step, $\Delta x = h = L/m$ is space step, $t_k = k\tau(k = 0,1,2,\cdots ,n)$, $x_i = ih(i = 0,1,2,\cdots , \quad m)$. For the time fractional derivative, we usually adopt following finite difference approximation

$$\frac{\partial^{\alpha} y(x_i,t_{k+1})}{\partial t^{\alpha}}$$
$$= \frac{1}{\Gamma(1-\alpha)} \sum_{j=0}^{k} \frac{y(x_i,t_{j+1}) - y(x_i,t_j)}{\tau} \int_{j\tau}^{(j+1)\tau} \frac{d\xi}{(t_{k+1}-\xi)^{\alpha}} + \tau$$
$$= \frac{\tau^{1-\alpha}}{\Gamma(2-\alpha)} \sum_{j=0}^{k} \frac{y(x_i,t_{k+1-j}) - y(x_i,t_{k-j})}{\tau} [(j+1)^{1-\alpha} -$$
$$j^{1-\alpha}] + O(\tau),$$
(11)

$$\frac{\partial y(x_i,t_{k+1})}{\partial x} = \frac{y(x_i,t_{k+1}) - y(x_{i-1},t_{k+1})}{h} + O(h).$$
(12)

For $\frac{\partial^{\beta} y(x,t)}{\partial x^{\beta}}$, by using Grünwald’s improved formula [4], we have that

$$\frac{\partial^{\beta} y(x_i,t_{k+1})}{\partial x^{\beta}} = \frac{1}{h^{\beta}} \sum_{j=0}^{i+1} a_j y(x_i - (j-1)h, t_{k+1}) + O(\tau + h),$$

(13)

where

$$a_0 = 1, a_1 = -\beta, a_j = (-1)^j \binom{\beta}{j},$$

$$\binom{\beta}{j} = \beta(\beta-1)\cdots(\beta - j + 1)/(j!), \ j = 1,2,3,\cdots .$$

Substituting Eqs. (11)-(13) into Eqs. (6)-(8), we obtain

$$\frac{\tau^{1-\alpha}}{\Gamma(2-\alpha)} \sum_{j=0}^{k} \frac{y(x_i,t_{k+1-j}) - y(x_i,t_{k-j})}{\tau} [(j+1)^{1-\alpha} - j^{1-\alpha}]$$
$$= \frac{u(x_i)}{h^{\beta}} \sum_{j=0}^{i+1} a_j y(x_i - (j-1)h, t_{k+1}) \text{ - } c(x_i) y(x_i, t_{k+1})$$

$$\text{-}v(x_i) \frac{y(x_i,t_{k+1}) \text{ - } y(x_{i\text{-}1},t_{k+1})}{h} + b(x_i,t_{k+1}) + O(\tau + h).$$

Let $u_i = u(x_i), v_i = v(x_i), c_i = c(x_i), \tilde{c}_i = \tau^{\alpha}\Gamma(2-\alpha) \geq 0,$

$b_i^k = b(x_i,t_k), \quad \sigma_j = (j+1)^{1-\alpha} - j^{1-\alpha}, \quad j = 0,1,2,\cdots ,n;$
$r_i =$
$\frac{u_i\tau^{\alpha}}{h^{\beta}}\Gamma(2-\alpha) \geq 0, p_i = \frac{v_i\tau^{\alpha}}{h}\Gamma(2-\alpha) \geq 0, g_1^k = g_1(t_k), g_2^k$
$= g_2(t_k)$, y_i^k represents numerical solution of Eqs. (6)-(8), then we have

$$\sum_{j=0}^{k} \sigma_j (y_i^{k+1-j} - y_i^{k-j}) = -p_i(y_i^{k+1} - y_{i-1}^{k+1}) + r_i \sum_{j=0}^{i+1} a_j y_{i-j+1}^{k+1}$$

$$-\tilde{c}_i \tilde{y}_i^{k+1} + \tilde{c}_i \tilde{b}_i^{k+1}.$$
(14)

Since local truncation error is $O(\tau + h)$, thus the difference scheme is consistent[10]. Eqs. (14) will be repalced by

When $k = 0$,

$$-r_i y_{i+1}^1 + (1 + p_i - r_i a_1 + \tilde{c}_i) \tilde{y}_i^1 - (p_i + r_i a_2) y_{i-1}^1 - r_i \sum_{j=3}^{i+1} a_j y_{i-j+1}^1$$

$$= y_i^0 + \tilde{c}_i \tilde{b}_i^1,$$
(15)

when $k > 0$,

$$-r_i y_{i+1}^{k+1} + (1 + p_i - r_i a_1 + \tilde{c}_i) \tilde{y}_i^{k+1} - (p_i + r_i a_2) y_{i-1}^{k+1} - r_i \sum_{j=3}^{i+1} a_j y_{i-j+1}^{k+1}$$

$$= y_i^k - \sum_{j=0}^{k} \sigma_j (y_i^{k+1-i} - y_i^{k-j}) = (2 - 2^{1-\alpha}) y_i^k + \sigma_k y_i^0 + \tilde{c}_i \tilde{b}_i^{k+1} -$$
$$\sum_{j=1}^{k-1} y_i^{k-j} [2(j+1)^{1-\alpha} - (j+2)^{1-\alpha} - j^{1-\alpha}]$$

$$= d_1 y_i^k - \sum_{j=1}^{k-1} y_i^{k-j} d_{j+1} + \sigma_k y_i^0 + \tilde{c}_i \tilde{b}_i^{k+1},$$
(16)

where $d_j = \sigma_{j-1} - \sigma_j$, $i = 1,2,\text{L}\ ,m-1; k = 1,2,\text{L}\ ,n-1.$

III. Stability and Convergence of The Difference Scheme

Lemma 2.1 For arbitrary real number a,b,c,d, we have

$$-|a|+|b|-|c|-|d| \le |-a+b-c-d|.$$

Proof. From $|b| = |-a+b-c-d+a+c+d|$

$$\le |-a+b-c-d|+|a|+|c|+|d|,$$

we obtain $-|a|+|b|-|c|-|d| \le |-a+b-c-d|$.

Lemma 2.2 (1) $a_j > 0 (j \ge 2)$. (2) For any positive integer N, we have $\sum_{j=0}^{N} a_j < 0$.

Proof. (1) Note that $a_j = (-1)^j \binom{\beta}{j}$ and $1 < \beta < 2$, we know that $a_j > 0 (j \ge 2)$.

(2) Since $(1+x)^\beta = \sum_{j=0}^{\infty} \binom{\beta}{j} x^j, x \in [-1,1]$, let $x = -1$, then $\sum_{j=0}^{\infty} a_j = 0$. From Lemma 2.2(1), we have that

$$\sum_{j=0}^{N} a_j = -\sum_{j=N+1}^{\infty} a_j < 0.$$

Lemma2.3 $(1)\sum_{j=1}^{n} d_j = 1-\sigma_n$; $(2) d_j > 0, \sigma_{j-1} > \sigma_j$.

Proof. (1)From $d_j = \sigma_{j-1} - \sigma_j, \sigma_j = (j+1)^{1-\alpha} - j^{1-\alpha}$, we easily know that $\sum_{j=1}^{n} d_j = 1-\sigma_n$.

(2) Let $h(x) = (x+1)^{1-\alpha} - x^{1-\alpha} (x \ge 1)$, then $h'(x) = (1-\alpha)[(x+1)^{-\alpha} - x^{-\alpha}] < 0$, so $h(x)$ is decreasing function, $d_j = \sigma_{j-1} - \sigma_j = h(j-1) - h(j) > 0$. Therefore, we have that $d_j > 0, \sigma_{j-1} > \sigma_j$.

A. Stability of the difference scheme

Theorem 2.1 Implicit difference schemes defined by Eqs. (15)-(16) are unconditionally steady for initial value[10].

Proof. Suppose that $\tilde{y}_i^k, y_i^k$ represent solutions of Eqs. (15)-(16) for initial value $f_1(x), f_2(x)$ respectively, and b_i^k is accurate vale, then error $\varepsilon_i^k = \tilde{y}_i^k - y_i^k$ satisfies

When $k = 0$,

$$-r_i\varepsilon_{i+1}^1 + (1+p_i - r_i a_1 + \tilde{c}_i)\tilde{\varepsilon}_i^1 - (p_i + r_i a_2)\varepsilon_{i-1}^1$$

$$-r_i\sum_{j=3}^{i+1} a_j \varepsilon_{i-j+1}^1 = \varepsilon_i^0,$$

when $k > 0$,

$$-r_i\varepsilon_{i+1}^{k+1} + (1+p_i - r_i a_1 + \tilde{c}_i)\tilde{\varepsilon}_i^{k+1} - (p_i + r_i a_2)\varepsilon_{i-1}^{k+1}$$

$$-r_i\sum_{j=3}^{i+1} a_j \varepsilon_{i-j+1}^{k+1} = d_1\varepsilon_i^k + \sum_{j=1}^{k-1} d_{j+1}\varepsilon_i^{k-j} + \sigma_k\varepsilon_i^0.$$

Let $E^k = (\varepsilon_1^k, \varepsilon_2^k, \text{L}\ , \varepsilon_{m-1}^k)'$, we prove $\|E^k\|_\infty \le \|E^0\|_\infty$ with mathematical induction in the following.

When $k = 1$, let $|\varepsilon_l^1| = \max_{1 \le i \le m-1} |\varepsilon_i^1|$. Note that $r_l > 0$, $p_l > 0$, $a_0 = 1, a_1 = -\beta < 0$, we have from Lemma 2.2 that

$$\|E^1\|_\infty = |\varepsilon_l^1| \le |\varepsilon_l^1| + p_l(|\varepsilon_l^1| - |\varepsilon_{l-1}^1|) - r_l\sum_{j=0}^{l+1} a_j |\varepsilon_l^1|$$

$$\le -r_l|\varepsilon_{l+1}^1| + (1+p_l - r_l a_1 + \tilde{c}_i)|\tilde{\varepsilon}_l^1|$$

$$-(p_l + r_l a_2)|\varepsilon_{l-1}^1| - r_l\sum_{j=3}^{l+1} a_j|\varepsilon_{l-j+1}^1|.$$

Note that $p_l + r_l a_2 > 0, r_l\sum_{j=3}^{l+1} a_j > 0, 1+p_l - r_l a_1 + \tilde{c}_i > 0$, and from Lemma 2.1, we further obtain that

$$|\varepsilon_l^1| \le |-r_l\varepsilon_{l+1}^1 + (1+p_l - r_l a_1 + \tilde{c}_i)\tilde{\varepsilon}_l^1 - (p_l + r_l a_2)\varepsilon_{l-1}^1$$

$$-r_l\sum_{j=3}^{l+1} a_j\varepsilon_{l-j+1}^1| = |\varepsilon_l^0| \le \|E^0\|_\infty.$$

Thus $\|E^1\|_\infty = |\varepsilon_l^1| \le \|E^0\|_\infty$.

Assume that we always have $\|E^k\|_\infty \le \|E^0\|_\infty$ when $k \le s$, let $|\varepsilon_l^{s+1}| = \max_{1 \le i \le m-1}|\varepsilon_i^{s+1}|$, then when $k = s+1$, we have

$$|\varepsilon_l^{s+1}| \le -r_l|\varepsilon_{l+1}^{s+1}| + (1+p_l - r_l a_1 + \tilde{c}_i)|\tilde{\varepsilon}_l^{s+1}|$$

$$-(p_l + r_l a_2)|\varepsilon_{l-1}^{s+1}| - r_l\sum_{j=3}^{l+1} a_j|\varepsilon_{l-j+1}^{s+1}|$$

$$\le |-r_l\varepsilon_{l+1}^{s+1} + (1+p_l - r_l a_1 + \tilde{c}_i)\tilde{\varepsilon}_l^{s+1}$$

$$-(p_l + r_l a_2)\varepsilon_{l-1}^{s+1} - r_l \sum_{j=3}^{l+1} a_j \varepsilon_{l-j+1}^{s+1} |$$

$$= | d_1 \varepsilon_l^s + \sum_{j=1}^{s-1} d_{j+1}\varepsilon_l^{s-j} + \sigma_s \varepsilon_l^0 |$$

$$\le d_1 \| E^s \|_\infty + \sum_{j=1}^{s-1} d_{j+1} \| E^{s-j} \|_\infty + \sigma_s \| E^0 \|_\infty$$

$$\le (d_1 + \sum_{j=1}^{s-1} d_{j+1} + \sigma_s) \| E^0 \|_\infty$$

$$= (1 - \sigma_s + \sigma_s) \| E^0 \|_\infty = \| E^0 \|_\infty .$$

Consequently, the desired result follows.

B. Convergence of the difference scheme

Suppose that $y(x_i, t_k)$ is exact solution of the differential equation at (x_i, t_k). Let $e_i^k = y(x_i, t_k) - y_i^k$, $e^k = (e_1^k, e_2^k, \cdots, e_{m-1}^k)'$, then $y_i^k = y(x_i, t_k) - e_i^k$, substituting it into Eqs. (15)-(16), and note $e^0 = 0$, we have that

When $k = 0$,

$$-r_i e_{i+1}^1 + (1 + p_i - r_i a_1 + \tilde{c}_i) e_i^1 - (p_i + r_i a_2) e_{i-1}^1$$

$$-r_i \sum_{j=3}^{i+1} a_j e_{i-j+1}^1 = R_i^1,$$

when $k > 0$,

$$-r_i e_{i+1}^{k+1} + (1 + p_i - r_i a_1 + \tilde{c}_i) e_i^{k+1} - (p_i + r_i a_2) e_{i-1}^{k+1}$$

$$-r_i \sum_{j=3}^{i+1} a_j e_{i-j+1}^{k+1} = d_1 e_i^k + \sum_{j=1}^{k-1} d_{j+1} e_i^{k-j} + R_i^{k+1},$$

where $| R_i^k | \le \lambda(\tau^{1+\alpha} + \tau^\alpha h), \lambda$ is a positive constant and it has nothing to do with h, τ.

Theorem 2.2 There is a constant $\lambda > 0$ such that

$$\| e^k \|_\infty \le \sigma_{k-1}^{-1} \lambda(\tau^{1+\alpha} + \tau^\alpha h), k = 1, 2, \cdots, n,$$

where λ has nothing to do with h, τ, and $\| e^k \|_\infty = \max_{1 \le i \le m-1} | e_i^k |$.

Proof. When $k = 1$, let $| e_l^1 | = \max_{1 \le i \le m-1} | e_i^1 |$, then $\| e^1 \|_\infty = | e_l^1 |$, we have from Lemma 2.1 that

$$| e_l^1 | \le -r_l | e_{l+1}^1 | + (1 + p_l - r_l a_1 + \tilde{c}_i) | e_l^1 |$$

$$-(p_l + r_l a_2) | e_{l-1}^1 | - r_l \sum_{j=3}^{l+1} a_j | e_{l-j+1}^1 |$$

$$\le | -r_l e_{l+1}^1 + (1 + p_l - r_l a_1 + \tilde{c}_i) e_l^1 - (p_l + r_l a_2) e_{l-1}^1$$

$$-r_l \sum_{j=3}^{l+1} a_j e_{l-j+1}^1 | = | R_l^1 |$$

$$\le \lambda(\tau^{1+\alpha} + \tau^\alpha h) = \sigma_0^{-1} \lambda(\tau^{1+\alpha} + \tau^\alpha h).$$

Assume that $\| e^s \|_\infty \le \sigma_{s-1}^{-1} \lambda(\tau^{1+\alpha} + \tau^\alpha h)$ when $k \le s$, let $| e_l^{s+1} | = \max_{1 \le i \le m-1} | e_i^{s+1} |$, then when $k = s + 1$, we have that

$$\| e^{s+1} \|_\infty = | e_l^{s+1} |$$

$$\le d_1 \| e^s \|_\infty + \sum_{j=1}^{s-1} d_{j+1} \| e^{s-j} \|_\infty + \lambda(\tau^{1+\alpha} + \tau^\alpha h)$$

$$\le (1 + d_1 \sigma_{s-1}^{-1} + d_2 \sigma_{s-2}^{-1} + \cdots + d_s \sigma_0^{-1}) \lambda(\tau^{1+\alpha} + \tau^\alpha h),$$

as in Lemma 2.3, we have found that $\sigma_j^{-1} \le \sigma_s^{-1} (j \le s)$, so we futher obtain

$$\| e^{s+1} \|_\infty \le \sigma_s^{-1} (\sum_{i=0}^{s-1} d_{i+1} + \sigma_s) \lambda(\tau^{1+\alpha} + \tau^\alpha h)$$

$$= \sigma_s^{-1} \lambda(\tau^{1+\alpha} + \tau^\alpha h).$$

The desired result follows.

Since $\lim_{k \to \infty} \frac{\sigma_k^{-1}}{k^\alpha} = \lim_{k \to \infty} \frac{1}{k^\alpha \sigma_k} = \lim_{k \to \infty} \frac{1}{k[(1 + 1/k)^{1-\alpha} - 1]} =$ $\frac{1}{1-\alpha}$, thus there is a constant $\gamma > 0$ such that

$$\| E^k \|_\infty \le k^\alpha \gamma(\tau^{1+\alpha} + \tau^\alpha h) = (k\tau)^\alpha \gamma(\tau + h).$$

Consequently, we can obtain the following result when $k\tau \le T$.

Theorem 2.3 Suppose that $y(x_i, t_k)$ denotes exact solution at (x_i, t_k), y_i^k is numerical solution of implicit difference scheme, then there exists a constant $\tilde{\gamma} = T^\alpha \gamma > 0$ such that

$$| y(x_i, t_k) - y_i^k | \le \tilde{\gamma}(\tau + h), \ 1 \le i \le m, 1 \le k \le n.$$

IV. Source Term Inversion

The inverse problem, which is composed of Eqs. (2)-(5), is to solve nonlinear operator equation

$$A[b(x)] = \psi(x).$$

Suppose that $y(b(x);x,t)$ denotes solution of Eqs. (2)-(4) for $b(x)$, $b_0(x)$ denotes a function near $b^*(x)$, where $b_0(x)= \sum_{i=1}^{n} k_i^0\psi_i(x) = K_0^{\mathrm{T}}\Psi(x)$, and $b^*(x)$ denotes exact solution of Eqs. (2)-(4), $b(x)\in K$, $\psi_1(x),\psi_2(x),\mathrm{L}$, are a group of primary functions on K , then a tiny disturbing quantity of $b_0(x)$ is

$$\delta b_0(x)=\sum_{i=1}^{n}\delta k_i^0\psi_i(x)=\delta K_0^{\mathrm{T}}\Psi(x),$$

(17)

where $\Psi(x)=(\psi_1(x),\psi_2(x),\mathrm{L}\ ,\psi_n(x))^T$, $K^T=(k_1,k_2,\mathrm{L}\ ,\ k_n)\in R^n$.

Assume that $y(b_1(x);x,t)$ denotes the solution of initial boundary value problem for $b_1(x)$, where $b_1(x)=b_0(x)$ $+\delta b_0(x)$, using Taylor formula, then we have[6]

$$y(b_0(x)+\delta b_0(x);x,t)$$

$$=y(b_0(x);x,t)+\nabla_{K_0}^T y(b_0(x);x,t)\cdot\delta K_0+o(\|\,\delta b_0(x)\,\|),$$

by using the Tikhonov regularization method, solving $b(x)$ is converted into δK_0 , and δK_0 can be determined by local minimum of the following function[9]

$$G[\delta K_0]{=}\|y(b_0(x)+\delta b_0(x);x,t)-y(x,T)\,\|^2_{L^2(\partial\Omega^{'}\times[0,T])}$$

$$+\alpha S[\delta K_0]\ {=}\|y(b_0(x);x,t)-y(x,T)+$$

$$\nabla_{K_0}^T y(b_0(x);x,t)\cdot\delta K_0\ \|^2_{L^2(\partial\Omega^{'}\times[0,T])}\ +\alpha S[\delta K_0],$$

(18)

where $x\in\partial\Omega^{'}\subset\partial\Omega$, α denotes regularization coefficient, $S[\delta K_0]$ denotes steady functional of δK_0.

Assume that there are discrete points $x_m(m=1,2,\mathrm{L}\ ,$ $M)$ on $\partial\Omega^{'}$, $S[\delta K_0]{=}\delta K_0^T\delta K_0$, then

$$G[\delta K_0]{=}\delta K_0^T A^T A\delta K_0+2\delta K_0^T A^T(P-Q)$$

$$+(P-Q)^T(P-Q)+\alpha\delta K_0^T\delta K_0.$$

(19)

It is easy to prove that solving the local minimum values of Eqs. (19) is equivalent to slove $(A^TA+\alpha I)\delta K_0=A^T$ $(Q-P)$, so we have

$$\delta K_0=(A^TA+\alpha I)^{-1}A^T(Q-P),$$

(20)

where

$$P=\begin{bmatrix} y(b_0(x);x_1,T) \\ y(b_0(x);x_2,T) \\ \mathrm{M} \\ y(b_0(x);x_M,T) \end{bmatrix},\qquad Q=\begin{bmatrix} y_T(x_1) \\ y_T(x_2) \\ \mathrm{M} \\ y_T(x_M) \end{bmatrix},$$

$$A=(a_{ij})_{M\times N},\ a_{ij}=\frac{\partial}{\partial k_j}y(b_0(x);x_i,T).$$

Substituting Eqs. (20) into Eqs. (17), we can obtain $\delta b_0(x)$ and a new approximation of the exact solution, $b_1(x)=b_0(x)+\delta b_0(x)$.

Repeating the above, until satisfies the precision requirement.

V. Numerical Experiments

In order to verify the effectiveness of the algorithm in the source term identification, we do the following numerical experiment[7]. For simplicity, we set part variables as follows

$$c(x)=0.05,u(x)=292,v(x)=365,L=4000,T=11,$$
$$m=20,n=11,\Delta=[0.01,0.01,0.01].$$

Where Δ is the increment of K when computing matrix A from Eqs. (20). Moreover, we always take polynomial function space as primary function space in the following computation, and setting initial boundary condition as follows

$$y_T(x)=0.057x+45.6,0\le x\le 4000,$$

$$g_1(t)=0.724t^2+45.6,0\le t\le 11,$$

$$g_2(t)=2.2t^2+273.4,0\le t\le 11.$$

Let $b(x)=1-x$ in the Eqs. (6)-(8), and substituting initial boundary condition, we can obtain $y(X,T)$ by sloving forward problem. And as the additional data, we can do inversion calculation by using the above algorithm. Let initial iteration vector $K_0=[1,1,1]$, and iterative termination condition $\delta b(x)<1e-4$, then we obtain inversion results under different regularization coefficient(see TABLE Ⅰ), let $theta=1e-3$, we get inversion results under different initial value(see TABLE Ⅱ).

TABLE Ⅰ. The Inversion Results Under Different Regularization Coefficient

θ	Iteration times	Results

1e-1	4	1.0000 -1.0000 0
1e-2	3	1.0000 -1.0000 0
1e-3	3	1.0000 -1.0000 0
1e-4	3	1.0000 -1.0000 0

TABLE Ⅱ. THE INVERSION RESULTS UNDER DIFFERENT INITIAL VALUE

K_0	Iteration times	Results
-100 -100 -100	4	1.0000 -1.0000 0
-10 -10 -10	3	1.0000 -1.0000 0
1 1 1	3	1.0000 -1.0000 0
10 10 10	3	1.0000 -1.0000 0
100 100 100	4	1.0000 -1.0000 0

To better simulate the errors generated by actual data, and verify the effectiveness of the algorithm, we choose the disturbance error $V^{\rho} = V(1+\xi\rho)$, where $\xi \in [-1,1]$, and $\rho > 0$ is error level.

According to the above algorithm, we do 8 times numerical experiments, and obtain the inversion results under different error level ρ (see TABLE Ⅲ). Besides, the comparison of inversion results and exact solution can see Figure 1 when ρ =0.01.

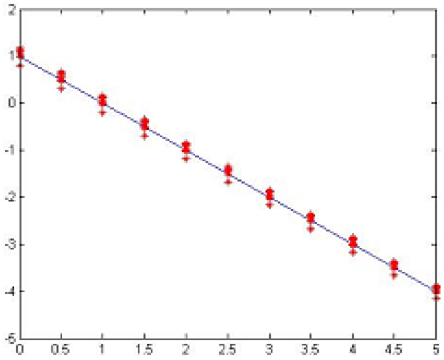


Figure 1. The comparison of inversion results and exact solutions

TABLE Ⅲ. THE INVERSION RESULTS UNDER DIFFERENT ERROR LEVEL

Times	ρ =0.01	ρ =0.05	ρ =0.1
1	0.9941 -0.9997 -0.0000	1.4988 -1.0243 0.0000	0.3838 -0.9700 -0.0000
2	1.1639 -1.0080 0.0000	0.3221 -0.9670 -0.0000	2.3115 -1.0639 0.0000
3	1.1098 -1.0053 0.0000	0.8026 -0.9903 -0.0000	-1.0527 -0.9000 -0.0000
4	0.9818 -0.9991 -0.0000	1.9119 -1.0444 0.0000	-0.5123 -0.9264 -0.0000
5	0.7984 -0.9902 -0.0000	1.8730 -1.0425 0.0000	-0.2448 -0.9394 -0.0000
6	1.1346 -1.0066 0.0000	0.8121 -0.9909 -0.0000	-0.2617 -0.9386 -0.0000
7	0.9768 -0.9989 -0.0000	1.8243 -1.0401 0.0000	1.4347 -1.0212 0.0000
8	1.0483 -1.0024 0.0000	0.0742 -0.9549 -0.0000	0.0459 -0.9535 -0.0000
mean value	0.9941 -1.0013 0.0000	1.1399 -1.0067 0.0000	0.2631 -0.9641 -0.0000

Through the above numerical experiment, we find that the inversion results and exact solution are almost the same, and this shows that the above algorithm is feasible and very effective.

REFERENCES

[1] Andreas Kirsch, An introduction to the mathematical theory of inverse problems, Springer, Karlsruhe, 1996.

[2] Gongsheng Li, Jinqing Liu, et al., A new gradient regularization algorithm for source term inversion in 1D solute transportation with final observations, Appl. Math. Comput, Vol.196,pp. 646-660, 2008.

[3] Yang Zhang, A finite method for fractional partial differential equation,Applied Mathematics and Computation, Vol.215, pp. 524-529, 2009.

[4] Meerschaert M M and Tadjeran C, Finite difference approximations for fractional advection-dispersion flow equations, J. Comput. Appl. Math., Vol.172,pp. 65-77, 2004.

[5] Podlubny I, Fractional differential equations, Academic Press, San Diego,1999.

[6] Jinqing Liu, Gongsheng Li and Yu Ma, Gradient—regulation method for determining a pollution source term in groundwater, Journal of Shandong University of Technology(Natural Science Edition), Vol. 21,No.2,pp.17-21,2007.

[7] Gongsheng Li, Yongji Tan and Xiaoqin Wang, Inverse problem method on determining magnitude of groundwater pollution sources, Mathematica Applicata, Vol.18, No.1, pp.92-98, 2005.

[8] Nazheng Sun, Inverse problem in groundwater modeling. Kluwer, Dordrecht, 1994.

[9] Tingyan Xiao, Shengen Yu and Yanfei Wang, Numerial solution of inverse problem, Beijing: Science Press, 2003.

[10] Wensheng Zhang, The finite difference method of partial differential equation in science calculation, Beijing: Higher Education Press, 2006.

e-Government Ethics : a Synergy of Computer Ethics, Information Ethics, and Cyber Ethics

Arief Ramadhan, Dana Indra Sensuse, Aniati Murni Arymurthy
Faculty of Computer Science
University of Indonesia
Depok, Indonesia

Abstract—**Ethics has become an important part in the interaction among humans being. This paper specifically discusses applied ethics as one type of ethics. There are three applied ethics that will be reviewed in this paper, i.e. computer ethics, information ethics, and cyber ethics. There are two aspects of the three applied ethics that were reviewed, i.e. their definition and the issues associated with them. The reviewing results of the three applied ethics are then used for defining e-Government ethics and formulating the issues of e-Government ethics. The e-Government ethics position, based on the previous three applied ethics, is also described in this paper. Computer ethics, information ethics and cyber ethics are considered as the foundations of e-Government ethics and several others applied ethics could enrich the e-Government ethics.**

Keywords- e-Government; Ethics; Applied Ethics; Computer Ethics; Information Ethics; Cyber Ethics

I. INTRODUCTION

Basically, the ethics regulates human behavior in doing something, whether someone doing the right thing or wrong thing. In determining whether someone doing is true or not, ethic is more concerned to the acceptability by his social environment. In this sense, ethics are social centric. An individual can not properly claim that his action is right ethically, unless their social environment consider it correct. This is consistent with what is stated in the [3], that ethic is relationship conduct pattern based on respect own rights and others against their environment.

Ethics is closely related to philosophy, so that several definitions of ethics would involve the word philosophy in it. As stated in [4], ethics is a branch of philosophy that is concerned with human conduct, more specifically the behavior of individuals in society. Other definition in [5] says that ethics is a branch of philosophy that deals with what is considered to be right and wrong. In [6], it is described that Ethics is a branch of philosophy that studies morals and values. In addition, another definition states that the field of ethics (or moral philosophy) involves systematizing, defending, and recommending concepts of right and wrong behavior [7].

There are two aspects in the definition of ethics: being able to determine what is right or wrong, good or bad and a commitment to doing what is right and good [4]. Ethics examines the rational justification for our moral judgments; it studies what is morally right or wrong, just or unjust [4]. Ethics are a subset of values: a value applies to things that are desired as well as what one ought to do, and can include such concepts as wealth, happiness, success, and fulfillment [4].

If we examine some various explanations above, it appears that ethics is closely related to morality. However, ethics can be not the same as morality. As hinted in [8], morality will be understood as the set of norms that guide our factual behavior whereas ethics is seen to be the theory and reflection of morality.

As stated in [7], philosophers today usually divide ethical theories into three general subject areas: metaethics, normative ethics, and applied ethics. Metaethics investigates where our ethical principles come from, and what they mean [7]. Metaethical answers to these questions focus on the issues of universal truths, the will of God, the role of reason in ethical judgments, and the meaning of ethical terms themselves [7]. When compared to normative ethics and applied ethics, the field of metaethics is the least precisely defined area of moral philosophy [7]. We may define metaethics as the study of the origin and meaning of ethical concepts [7].

Unlike the metaethics, normative ethics takes on a more practical task, which is to arrive at moral standards that regulate right and wrong conduct [7]. This may involve articulating the good habits that we should acquire, the duties that we should follow, or the consequences of our behavior on others [7]. In a sense, it is a search for an ideal litmus test of proper behavior [7]. The Golden Rule is a classic example of a normative principle: We should do to others what we would want others to do to us [7].

The key assumption in normative ethics is that there is only one ultimate criterion of moral conduct, whether it is a single rule or a set of principles [7]. Three strategies that are associated with normative ethics are also revealed in the [7], i.e. virtue theories, duty theories, and consequentialist theories.

Applied ethics can be classified into several types. This division is generally adapted to the needs of the social environment. Fieser in the [7], states that Applied ethics is the branch of ethics which consists of the analysis of specific, controversial moral issues. Other statements is revealed by Kaddu in [4], that ethics leads to a set of rules of conduct for specific situations; basic ethical principles guide the development of standards for specific professions and groups. What was put forward by Kaddu is highly relevant to applied ethics.

This paper will attempt to define what the meaning of e-Government ethics. The definition will be associated with the applied ethics. There are three applied ethics, in the world of computers science, which will be used as a reference, i.e. computer ethics, information ethics, and cyber ethics. Therefore, this paper will also try to discuss these three applied ethics from several perspectives. Some issues related to the three applied ethics will also be identified.

In addition this paper will also describe the position of the e-Government ethics against computer ethics, information ethics and cyber ethics. A diagram of this concept will be used to describe the position of the e-Government ethics.

II. A Review on Computer Ethics Definition and Issues

At this time, there are several emerging applied ethics, such as environment ethics, media ethics, etc. Several applied ethics that related to computer science world is computer ethics, information ethics and cyber ethics.

As the name implies, computer ethics is closely related to the use of computers by humans. We suggest that there are two things in the computer ethics that can be observed, i.e. whether the computer is used to do the right thing or the computers are used correctly.

In [9], it is revealed that computer ethics is not simply ethics rotely applied to computing. Computer ethics has two parts: (i) the analysis of the nature and social impact of computer technology and (ii) the corresponding formulation and justification of policies for the ethical use of such technology [9].

Computer ethics is a standard for computer use, signifying the prevention of copyright infringement, such as the reproduction of software, invasion of privacy, and circulation of objectionable material [10]. Computer ethics is made to research about security and it`s beneficial aspects [11]. Computer ethics is also used to refer to professional ethics for computer professionals such as ethical codes of conduct that can be used as guidelines for an ethical case [12]. Lee and Chan, in [13], suggest that the work of computer ethics is not to create a new system of ethics but rather to apply traditional ethics and to extend them to cover situations that are attributed to computers.

Other interesting opinions related to the computer ethics can be seen in the [14]. The paper has presented computer ethics as neither a list of ethical principles to obey, nor a technology deprived of certain values while implementing those principles. Thus, computer ethics urges scholars to revisit computer technology and its values [14]. Even though computer ethics is a field related to and in between science and ethics, it is a unique and holistic discipline providing principles for understanding, conceptualization and computer technology use [14].

Brey in [15] suggests that there exist two parts of computer ethics, i.e. mainstream computer ethics and disclosive computer ethics. We consider, in accordance with our focus, mainstream computer ethics is relevant to our discussion.

In mainstream computer ethics, a typical study begins by identifying a morally controversial practice, like software theft, hacking, electronic monitoring, or internet pornography [15]. Next, the practice is described and analyzed in descriptive terms, and finally, moral principles and judgments are applied to it and moral deliberation takes place, resulting in a moral evaluation, and optionally, a set of policy recommendations [15]. There are three features of mainstream computer ethics proposed in [15], i.e. (1) Mainstream computer ethics focuses on existing moral controversies; (2) Its focus is on practices, i.e. the individual or collective behavior of persons, and it aims to evaluate and devise policies for these practices; (3) Its focus usually is on the use of computer technology, as opposed to, e.g., its design or advertisement.

Moor suggests that what is special about computer ethics is that it has a continually large number of evolving situations which are difficult to conceptualize clearly and for which it is hard to find justified ethical policies [9]. In summary, what is unique about computer ethics is computing technology itself, and what makes computer ethics different as a field of ethics is their scope, depth, and novelty of ethical situations for which conceptual revisions and policy adjustments are required [9].

From some of the above explanations, it can be concluded that there are several issues related to computer ethics, i.e. reproduction of software, invasion of privacy, circulation of objectionable material, and security. Several other issues are software theft, hacking, electronic monitoring, and internet pornography. Some of these issues can also appear in information ethics or cyber ethics.

III. A Review on Information Ethics Definition and Issues

In simple terms, information ethics can be interpreted as ethics in the using, accessing and disseminating the information. In this case, the information is used for the right things, the information accessed in the right way, and the information is delivered correctly to the hand who have the rights.

Information ethics has been developed since the 1980s, encompassing areas such as computer ethics and global information ethics [10]. Capurro and Britz, in [16], stated that information ethics is not only about norms, but also about our critical reflection on the visions and options for better lives in the digital age.

Information ethics is an open space of reflection where commonalities and differences, theoretical as well as practical, can be discussed without the immediate pressure of decision making [16]. Information ethics is the new ecological ethics for the information environment [17]. Information ethics is essentially concerned with the question of who should have access to what information [18]. Information ethics relates to questions of ethics in terms of information or an information-oriented society [10]. This includes the standard for judging behavior of an individual or a member of community and classifying these as moral or immoral [10].

From some of the above explanations, it can be seen that information ethics is closely related to environmental and social. As revealed in [19], that information ethics is one aspect of a much larger philosophy known as social ethics.

It is revealed in [17], that we have to fight any kind of destruction, corruption, pollution, depletion (marked reduction in quantity, content, quality, or value) or unjustified closure of the infosphere, what shall be referred to here as information entropy. The ethical use of ICT and the sustainable development of an equitable information society need a safe and public infosphere for all, where communication and collaboration can flourish, coherently with the application of civil rights, legal requirements and the fundamental freedoms in the media [17].

Information ethics in the future should be a discipline that carries out four functions, i.e. : (1) information ethics is prescriptive ethics; (2) information ethics is preventive ethics; (3) information ethics is transformative ethics; and (4) information ethics must be universally global ethics, not one or the other, but must consist of both global and local disciplines [10]. Furthermore, in [10] also stated that for the proper use of information in an information society, the education relating to information ethics may present four goals, i.e. (1) respect for others must be cultivated; (2) although sharing beneficial information is welcome, other people's intellectual property right must not be infringed; (3) various forms of information will be used productively; and (4) telecommunications and the Internet will be used for acceptable time periods so that it does not harm actual life.

There are several issues related to information ethics. Several issues are emerged with different names or from different sources, but its essence remains the same. One of the most important topics in information technology ethics is privacy [20]. Fallis, in [18], saying that the core issues of information ethics include intellectual freedom, equitable access to information, information privacy, and intellectual property. In addition, in [19], it is also stated the two major issues of information technology, i.e. the conflict between observing others' privacy, and the simultaneous pursuit of individual freedom and autonomy. Another issue related to information ethics is responsibility and accountability. This is in line with what is also revealed in [19], that information ethics deals with the moral conduct of information-users based on their responsibility and their accountability.

IV. A Review on Cyber Ethics Definition and Issues

"Cyber" is a prefix used to describe people, things, or ideas that are connected with computers and the internet [21]. Therefore, cyber ethics is closely related to the development of internet technology, so that some definition of cyber ethics will include the internet or online terms in it. Indeed, some sources call cyber ethics with internet ethics. In this paper, the term cyber-ethics and internet ethics can interchangeable with each other

Cyber ethics is really about social responsibility in cyberspace [22]. As stated in [10], Cyber ethics is a system of standards that prescribe morality and immorality in cyberspace, signifying the preservation of freedom of expression, intellectual property, and privacy. Other definition of cyber ethics can be found in [23], that is cyber ethics is the discipline dealing with what is good and bad, and with moral duty and obligation as they pertain to online environments and digital media.

Lin, in [24], call cyber ethics as an internet ethics. In the paper, Lin stated that the right or wrong about the utility of internet by mankind can be called internet ethics. The utility of internet by mankind includes interpersonal communication, information's delivery, research, storage and so on [24].

It could be argued, that all of ethics, which applies in the computer ethics and information ethics can also apply on cyber ethics, but only focused on its application to the internet. As stated in the [25], the term of internet ethics can be thought as a special extention of computer ethics, but the main difference between them is that internet ethics is dealing with the behaviors performed in internet.

The challenge for cyber ethics is to discuss principles of morality that can guide human action so that people are empowered to establish a sustainable, participatory global information society [26]. Cyber ethics can discuss real possibilities of development of the information society and criticize ideologies that portray the information society in uncritical and one-dimensional ways [26]. By e-mail or newsgroup, any sort of opinions and thoughts can be spread all over the world. On the one hand, it can help people communicate, express opinions and thoughts, and get responses from other people fast; on the other hand, it also may be misused, such as quite a few bothering ads, fraud letters, nonsense articles interfering seriously with other people's chances to get useful information [24].

Some of cyber ethics issues raised in [23], i.e. plagiarism, copyright, hacking, fair use, file sharing, online etiquette protocols, posting incorrect/inaccurate information, cyberbullying, stealing or pirating software, music, and videos, online gambling, gaming, and internet addiction. Several other current cyber ethics issues are raised in [27], i.e. privacy, security, electronic monitoring of employees, collection and use of personal information on consumers, and identity theft.

V. The Definition, Position and Issues of E-Government Ethics

Simply, e-Government ethics can be defined as ethics in the use of e-Government system, either to insert or update content into the system or to get content from the system. However, to give more in-depth understanding, we will try to explain it further by reviewing the definition of e-Government and see how e-Government system is implemented.

e-Government is the use of Information Technology (IT) by public sector organizations [1]. Other definition of e-Government is public sector use of the internet and other digital devices to deliver services, information, and democracy itself [2].

e-Government is also an information system [1]. So, it can be said that several theories about the information system can be applied in e-Government. However, e-Government is different from ordinary information system that is generally targeting the private sector. The main orientation of e-Government is the accessibility of information by the public, rather than financial income [1].

Because the target of e-Government is the public sector, then the e-Government systems are generally built based on the web technology. This technology is used because it has ability to reach people quickly and widely. This also implies that the users of e-Government systems will generally using computer in accessing the system.

e-Government has one of the characteristics of postmodernism, that is the social construct [30]. This shows the presence of interaction between human beings in e-Government. The interaction of course need a set of rules to regulate it. One set of rules that can be applied is ethics.

Based on the above definition and description regarding to e-Government, then it can be said, that in e-Government, it could apply three applied ethics, i.e. computer ethics, information ethics and cyber ethics. But, beside these three applied ethics, in e-Government, it also could apply some others applied ethics.

e-Government is not simply about information technology or website, but there are also some aspects of management in it. As revealed in [1], that e-Government are socio-technical and there are two aspects in e-Government, i.e. the technical aspects (technology) and managerial aspects. This means, that there is other ethics that can be applied, that is management ethics.

Moreover, one of the actors involved in e-Government is the business. So, the business ethics can also be applied in e-Government.

There is another aspect in e-Government, that is the aspect of the object being observed. For example, in [28], it is proposed a new paradigm in e-Government called e-Livestock. From the definition of e-Livestock, it can be seen that the object related to e-Livestock is the animal, i.e. cows or buffaloes. In this case, in addition to several ethics already discussed above, there are other ethics that can be related to e-Government, that is the ethics of animal treatment.

Computer ethics, information ethics, and cyber ethics can be said as the foundations of e-Government ethics, and there are another applied ethics as the complement of the e-Government ethics. Fig. 1 shows the position of e-Government ethics in relation to the other applied ethics.

Based on Fig. 1, it can also be concluded that some issues related to computer ethics, information ethics, and cyber ethics, which has been mentioned previously, could also become an issue in e-Government ethics. But, of course, it could be added with other issues, such as the issue of sensitivity of the information as revealed in [29], or the issue of trustworthiness of the content of e-Government system.

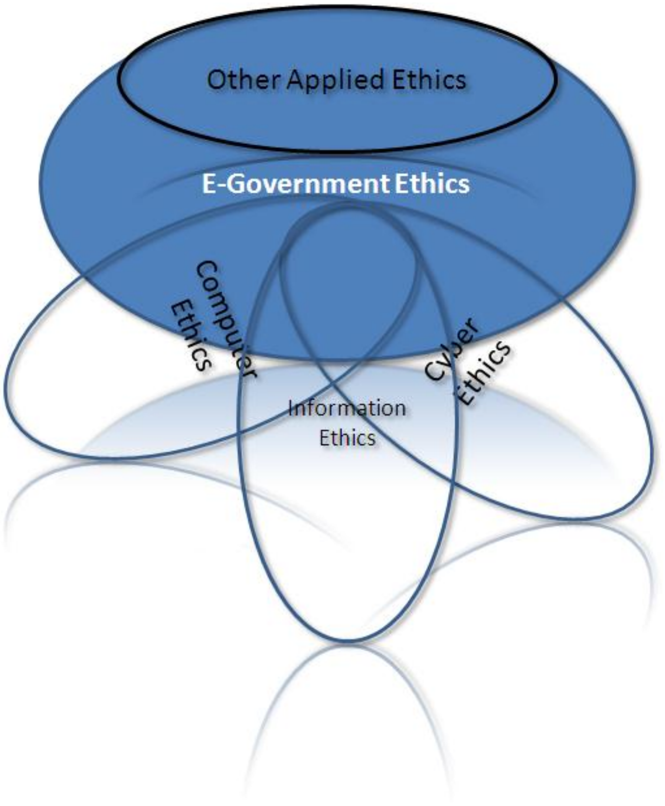


Figure 1. e-Government ethics position related to computer ethics, information ethics, cyber ethics and other applied ethics.

VI. Conclusions

This paper has summarized the definitions and several issues related to computer ethics, information ethics and cyber ethics. The three applied ethics turns out to be a foundation for e-Government ethics.

The definition of the e-Government ethics has been given in this paper. The position of e-Government ethics among computer ethics, information ethics, cyber ethics, and other applied ethics (such as the ethics of management, business, object treatment, etc.) is also described in this paper.

This paper can be the starting point of research about e-Government ethics. In the future, there could be many other applied ethics and issues that can be identified and added to the e-Government ethics presented in this paper.

References

[1] R. Heeks, Implementing and Managing eGovernment An International Text, London, England : SAGE Publications, 2006.

[2] D. M. West, Digital Government Technology and Public Sector Performance, New Jersey, USA : Princeton University Press, 2005.

[3] A. Shalbaf, “A view of Problems and Practical Pattern of Promotion of Ethics in Educational Organizatons,” Iranian Journal of Ethics in Science and Technology, vol. 4, no. 1, 2009.

[4] S. B. Kaddu, “Information Ethics: a student’s perspective,” International Review of Information Ethics, vol. 7, 2007.

[5] U. Averweg and G. Erwin, "Towards a Code of Cyberethics for a Municipality in South Africa," in Proceedings of the Fifth International Conference on Electronic Business, 2005, pp. 522-527.

[6] R. Yucel, H. Elibol, and O. Dagdelen, "Globalization and International Marketing Ethics Problems," International Research Journal of Finance and Economics, issue. 26, pp. 93-104, 2009.

[7] J. Fieser, "Ethics", Internet Encyclopedia of Philosophy, A Peer-Reviewed Academic Resources. [Online]. Available: http://www.iep.utm.edu/ethics/, 2009.

[8] B. C. Stahl, "Information, Ethics, and Computers: The Problem of Autonomous Moral Agents," Minds and Machines, vol. 14, pp. 67-83, 2004.

[9] J. H. Moor, "Reason, Relativity, and Responsibility in Computer Ethics," Computers and Society, pp. 14-21, March 1998.

[10] H. Ki and S. Ahn, "A Study on the Methodology of Information Ethics Education in Youth," International Journal of Computer Science and Network Security, vol. 6, no. 6, pp. 91-100, 2006.

[11] M. Namayandeh and H. Taherdoost, "Review Paper on Computer Ethics and Related Research Models," Journal of Open Problems in Science and Engineering, vol. 1, no. 1, pp 42-47, 2009.

[12] R. Heersmink, J. V. D. Hoven, N. J. V. Eck, and J. V. D. Berg, "Bibliometric Mapping of Computer and Information Ethics," CWTS Working Paper Series, p. 13, 2010.

[13] W. W. Lee and A. K. K. Chan, "Computer Ethics: an Argument for Rethinking Business Ethics", in The 2nd World Business Ethics Forum: Rethinking the Value of Business Ethics, 2008.

[14] A. Kuzu, "Problems Related to Computer Ethics: Origins of The Problems and Suggested Solutions," The Turkish Online Journal of Educational Technology, vol. 8, issue. 2, pp. 91-110, 2009.

[15] P. Brey, "Disclosive Computer Ethics: The Exposure and Evaluation of Embedded Normativity in Computer Technology," Computers and Society, vol. 30, no. 4, pp. 10-16, 2000.

[16] R. Capurro and J. B. Britz, "In search of a code of global information ethics: The road travelled and new horizons," Ethical Space: The International Journal of Communication Ethics, vol. 7, no. 2/3, pp. 28-36, 2010.

[17] L. Floridi, "Information Ethics: An Environmental Approach to the Digital Divide," Philosophy in the Contemporary World, vol. 9, no. 1, 2001.

[18] D. Fallis, "Information ethics for twenty-first century library professionals," Library Hi Tech, vol. 25, no. 1, pp. 23-36, 2007.

[19] C. P. Chuang and J. C. Chen, "Issues in Information Ethics and Educational Policies for the Coming Age," Journal of Industrial Technology, vol. 15, no. 4, 1999.

[20] A. R. Peslak, "Current Key Privacy Factors: Development and Analysis," Journal of Information Technology Impact, vol. 6, no. 3, pp. 171-186, 2006.

[21] F. L. Wilczenski and S. M. Coomey, "Cyber-Communication: Finding Its Place in School Counseling Practice, Education,and Professional Development," ASCA, pp 327-331, 9:4 April 2006.

[22] S. Mahfood, A. Astuto, R. Olliges, and B. Suits, "Cyberethics Social Ethics Teaching in Educational Technology Programs," Communication Research Trends, vol. 24, no. 4, pp. 3-43, 2005.

[23] D. Pruitt-Mentle, "2008 National Cyberethics, Cybersafety, Cybersecurity Baseline Study", Educational Technology Policy Research and Outreach (ETPRO), National Cyber Security Alliance (NCSA), October 2008.

[24] J. X. Lin, "Educational Enlightenments from Internet Ethics Issues," Journal of Information, Technology and Society, 2003(2), pp. 65-72.

[25] F. Odabasi and E. B. Kuzu, "A Proposal For Ethics Training In Internet Cafes," in Proceedings of the 7th WSEAS International Conference on Education and Educational Technology, pp. 141-144, 2008.

[26] C. Fuchs, R. M. Bichler, and C. Raffl, "Cyberethics and Co-operation in the Information Society," Sci Eng Ethics, Vol. 15, pp. 447-466, 2009.

[27] T. A. Kraft and J. Carlisle, "Computer Ethics: A Slow Fade from Black and White to Shades of Gray," in Proceedings of Information Systems Educators Conference, 2010.

[28] A. Ramadhan and D. I. Sensuse, "e-Livestock as a New Paradigm in e-Government," in Proceedings of the 3rd International Conference on Electrical Engineering and Informatics (ICEEI 2011), IEEE Press, vol. 1, 2011.

[29] Z. Fang, "E-Government in Digital Era: Concept, Practice, and Development," International Journal of The Computer, The Internet and Management, vol. 10, no. 2, pp. 1-22, 2002.

[30] A. Ramadhan, D. I. Sensuse, and A. M. Arymurthy, "Postmodernism in e-Government," International Journal of Computer Science Issues, vol. 8, issue. 4, no. 1, pp. 623-629, July 2011.

AUTHORS PROFILE

Arief Ramadhan. B.Sc in Computer Science (Bogor Agricultural University, Indonesia, 2005), M.Sc in Computer Science (Bogor Agricultural University, Indonesia, 2010), Ph.D Student in Computer Science (University of Indonesia), Research Assisstant at University of Indonesia. Member of e-Government Lab at University of Indonesia.

Dana Indra Sensuse. B.Sc in Geology (Bogor Agricultural University, Indonesia, 1985), M.Sc in Library and Information Studies (Dalhousie University, Halifax, Canada, 1994), Ph.D in Information Studies (Toronto University, Canada, 2004), Lecturer at University of Indonesia, Head of e-Government Lab at University of Indonesia.

Aniati Murni Arymurthy. B.Sc in Electrical Engineering (University of Indonesia, 1973), M.Sc in Computer and Information Sciences (Ohio State Univ., Ohio, USA, 1981), Ph.D in Computer Science (University of Indonesia, 1997), Professor at Faculty of Computer Science in University of Indonesia, Lecturer at University of Indonesia, Head of Pattern Rec, Image Proc, and CBIR Labs at University of Indonesia.

CO_2 Concentration Change Detection in Time and Space Domains by Means of Wavelet Analysis of MRA: Multi-Resolution Analysis

Kohei Arai
Department of Information Science,
Graduate School of Science and Engineering
Saga University
Saga city, Japan

***Abstract*—A method for change detection in time and space domains based on wavelet MRA: Multi-Resolution Analysis is proposed. Measuring stations for carbon dioxide concentration are sparsely situated. Also the measuring stations monitor the carbon dioxide concentration in an irregular basis so that some interpolation techniques are required for creation of carbon dioxide concentration with a regular interval in both time and space domains. After all, time and space of MRA is applied to the interpolated carbon dioxide concentration then reconstruction is done without low frequency of LLL component. Thus relatively large change locations and time periods are detected. Through an experiment with 11 years of carbon dioxide concentration data starting from 1990 provided by WDCGG: World Data Center for Greenhouse Gasses, it is found that there exists seasonal change and relatively large changes are occurred in El Nino years. It is also found that the carbon dioxide is concentrated in European continent.**

Keywords- wavelet analysis; carbon di-oxide distribution; change detection.

I. INTRODUCTION

As the results from the investigations in the pacific, Atlantic and Indian oceans for one decade starting from 1989 and ended in 1998, it was fund that 48% of carbon dioxide of generated from the human activity of fossil derived fuels consumptions and is put into the atmosphere absorbed by the oceans [1], [2]. Not only global carbon dioxide distributions but also regional distributions have being measured and estimated [3].

The carbon dioxide distributions in the atmosphere are measured and estimated at the fixed observation stations irregularly. It is getting much important to identify areas and time periods of which carbon dioxide amount changes in time and space domains through the investigations and analyzing the incomplete measured data for estimation of contiguous series of carbon dioxide distributions with interpolations in time and space domains. In order to detect the carbon dioxide distribution changes, wavelet based Multi Resolution Analysis (MRA) with appropriate parameters, base function and the level of MRA which is corresponding to the frequency component of the changes is proposed. Through experiments with the proposed MRA based method, relatively great changes are found and are highly related to the El Nino. Daubechies and Haar base function for wavelet based MRA are tried with the different levels for determination of most appropriate base function and levels which corresponds to frequency component it may concern.

The following chapter describes the proposed method followed by some experiments with the 10 years of carbon dioxide distribution data. Then, finally, conclusions and some discussions are followed.

II. PROPOSED METHOD

A. Interpolation Method

The carbon dioxide measurement stations are sparsely situated and collect carbon dioxide distribution (73 of measuring stations are situated in the world) in the atmosphere irregularly. Therefore interpolation methods are highly required in particular for space domain. The interpolation method used here can be represented with the following equations,

$$Z = \frac{\sum_{i=1}^{n} a_i z_i}{\sum_{i=1}^{n} a_i} \qquad (1)$$

$$a_i = \frac{1}{d_i^p} \qquad (2)$$

where z, z_i ,$a_{i,}$ d_i denote carbon dioxide concentration at the desired location, carbon dioxide concentration at the relatively closed measurement stations to the desired location, weighting coefficients, and the distance between the desired location and the closed measurement stations, respectively. Fig. 1 shows the definition of the desired location (black circle) and three of the closed measurement stations (red circles). Weighting coefficients are inversely proportional to the distance.

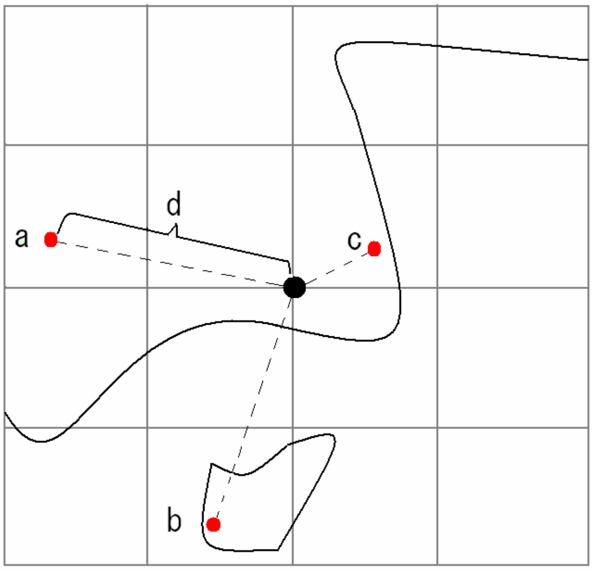


Figure 1 Definition of the desired location (black circle) and three of the closed measurement stations (red circles).

B. *Wavelet Based Multi Resolution Analysis(MRA)*

The locations and time periods representing relatively large changes of carbon dioxide distribution in time and space domains can be found through wavelet based MRA method. Namely, MRA is applied to the space and time series of carbon dioxide distribution of measured data and then reconstruct original data without low frequency component. Thus relatively large changes are corresponding to the comparatively high frequency components so that relatively large changes are detected.

III. Experimets

A. *The Data Used*

WDCGG (World Data Center for Greenhouse Gasses) of Japanese Meteorological Agency provides the measured carbon dioxide distribution data [5]. WDCGG was established to collect and distribute the data related to greenhouse gasses and the other related gasses in the atmosphere and the oceans as well as relating to global warming. They collect the data from the observation and measuring stations networks of the world atmospheric monitoring program and NOAA: National Oceanic and Atmospheric Administration. A small portion of data is shown in Table 1 and Fig. 2.

Fig. 2 shows the monthly trend of carbon oxide concentrations measured at 53:20N, 9:54W during from 1990 to 2005. It shows gradually increasing of the carbon dioxide concentrations as well as seasonal changes. The trend of the carbon dioxide concentration is almost same in the world. At least, the trend measured at the 73 of observation measurement stations show a similar trend of the trend of Fig. 2.

These data are a small portion of data. Other than these, there are a plenty of data which are provided by the stations and agencies.

Table 1 shows the monthly data of carbon dioxide in unit of ppm with the header information which includes location of the station, station name, data provider name etc. This format is common to all the measurement and observation stations.

Interpolated and re-configured data as one degree of meshed data is generated based on the proposed method. One of the examples of the re-configured data is shown in Fig. 3.

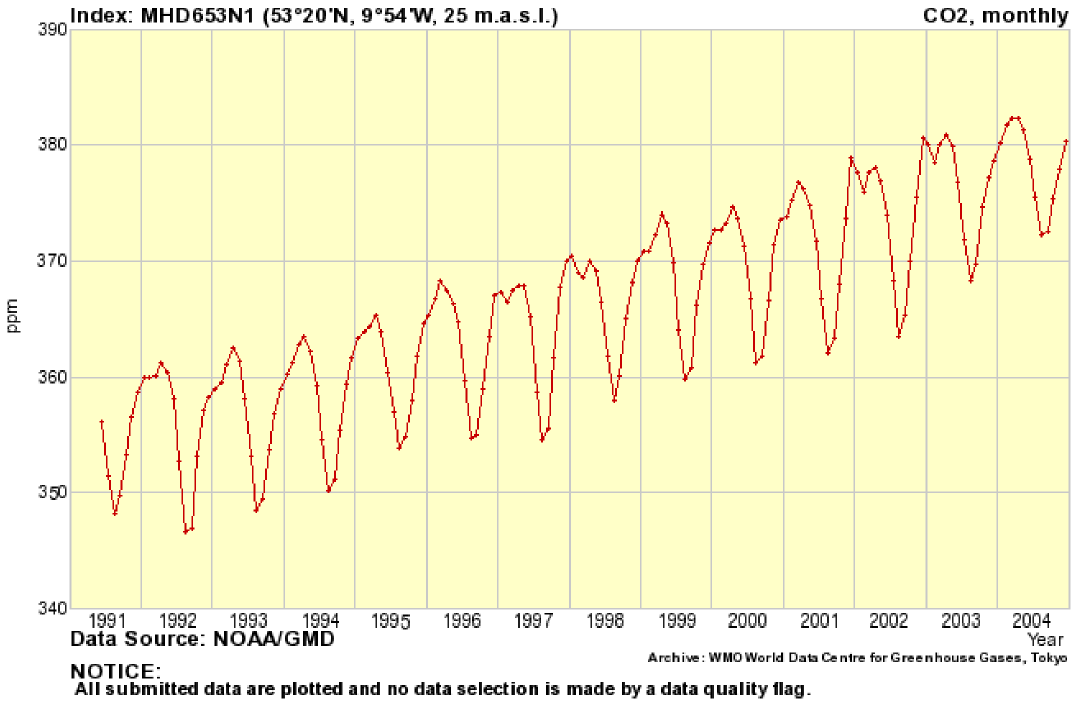


Figure 2 A small portion of measured data of the time series of carbon dioxide at 53:20N, 9:54W during from 1990 to 2005.

Warm colors mean densely concentration of carbon dioxide while cool colored areas show thinner carbon dioxide concentrations. Carbon dioxide concentration in the northern hemisphere is greater than that in the southern hemisphere. In this case of the carbon dioxide in January 1995, carbon dioxide concentration in European countries areas is densely situated, in particular, in comparison to the other areas. There are some artificial shapes of the densely distributed carbon dioxide concentration in the European countries areas. 1995 is not El Nino year so that not so warm sea surface temperature is not observed in the El Nino offshore area of Peru.

It is just an example. Another example of 1 degree mesh data of carbon dioxide distribution trend for 11 years starting from January 1990 is shown in Fig. 4. The car5bon dioxide distribution is shrinking in the horizontal direction, in particular, in order to show the trend much clearly. The black arrows in the bottom indicate El Nino years, 1990, 1993, and 1997-1998. As shown in Fig. 4, warm sea surface temperature is observed in the El Nino offshore areas coincidentally to the well known El Nino years.

B. Extraction of Spatial Features

In order to demonstrate a usefulness of the proposed method for extraction of spatial features, 2D wavelet transformation is applied to the spatial distribution data of carbon dioxide which is measured in April 1994. Fig. 5 (a) shows the original distribution while Fig. 5 (b), (c), and (d) shows the resultant images after the wavelet transformations.

TABLE I. SMALL PORTION OF MEASURED DATA OF THE TIME SERIES OF THE CARBON DIOXIDE AT MINAMI-TORISHIMA ISLAND, JAPAN.

```
REM01  EXPLANATORY 25-LINE FOR THE
MONTHLY DATA
REM02
REM03  STATION: Minamitorishima
REM04  CATEGORY: Global
REM05  COUNTRY/TERRITORY: Japan
REM06  SUBMITTED BY: JMA
REM07  LATITUDE:    24 18' N
REM08  LONGITUDE: 153 58' E
REM09  ALTITUDE:        8 m
REM10
REM24 M      CO2   V   ND     SD   F
REM25        ppm
1993 01 -9999999 2F9    0 -99999   2
1993 02   360.98 018   85  0.684   3
1993 03   360.57 012  570  0.942   3
1993 04   361.66 012  574  1.247   3
1993 05   360.78 012  526  0.998   3
1993 06   359.58 013  471  0.656   3
1993 07   357.69 011  630  1.114   3
1993 08   355.53 013  507  1.240   3
1993 09   353.71 011  635  1.111   3
1993 10   354.76 010  690  1.059   3
1993 11   355.91 014  404  0.802   3
1993 12   357.84 012  580  1.043   3
1994 01   359.67 011  648  0.816   3
1994 02   361.09 012  531  0.854   3
                             .
                             .
                             .
```

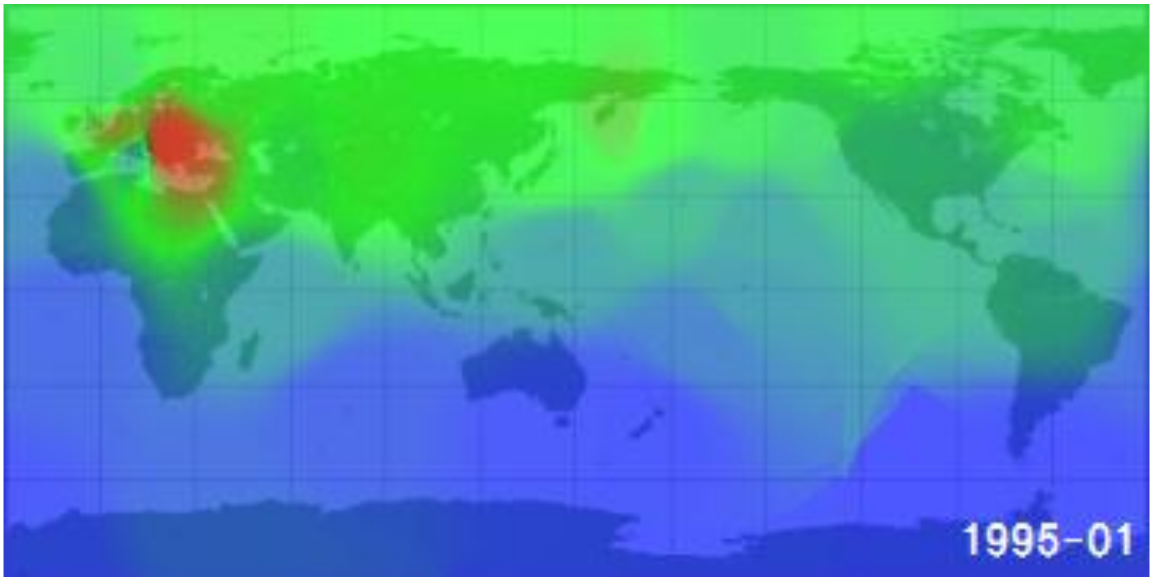


Figure 3 One degree meshed data of carbon dioxide distribution for January 1995.

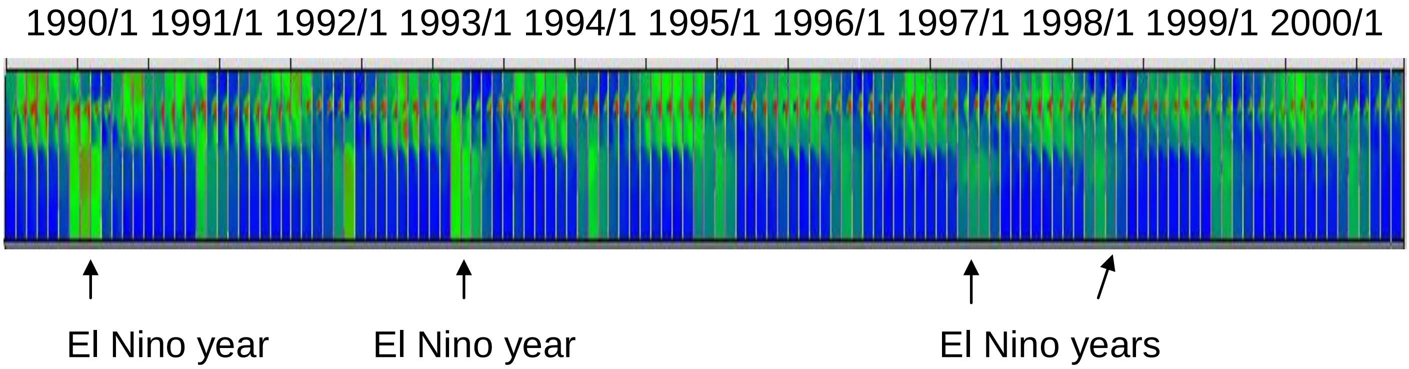
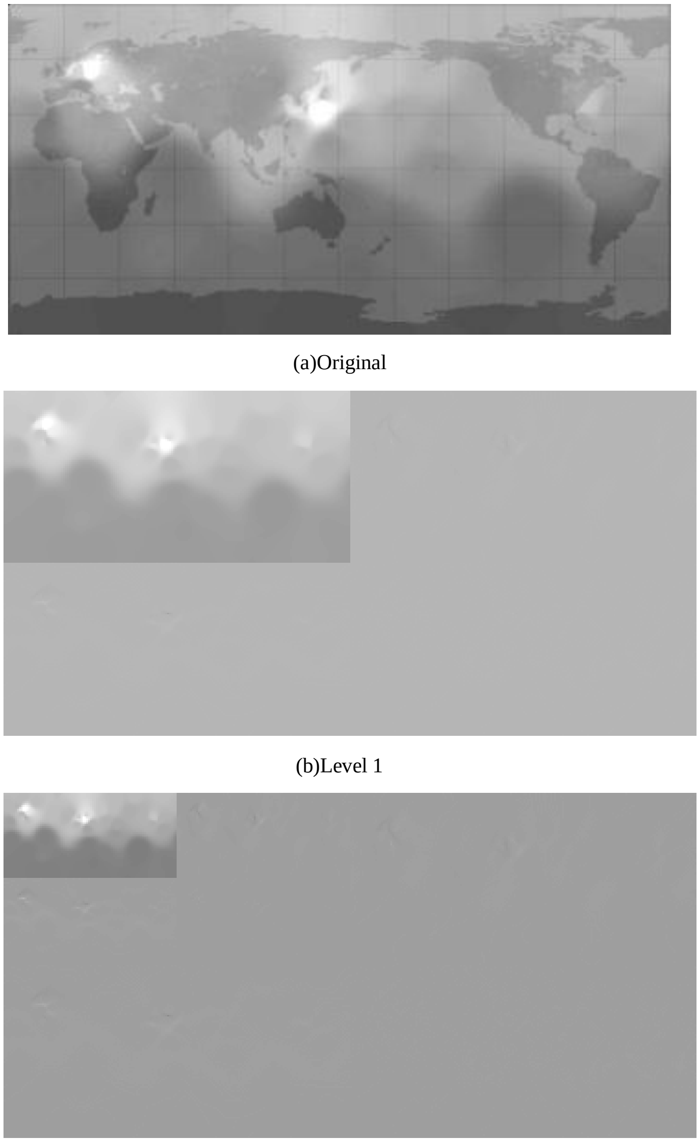


Figure 4 One degree mesh data of time series of carbon dioxide distribution trend for 11 years starting from January 1990.



(a)Original

(b)Level 1

(c)Level 2

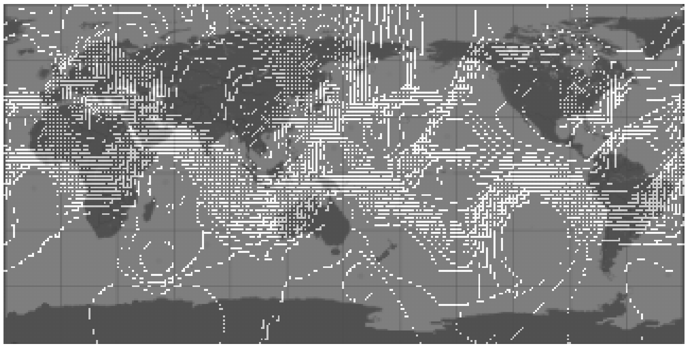
(d)Level 4

Figure 5 Spatial variability detection of carbon dioxide distributions in April 1994.

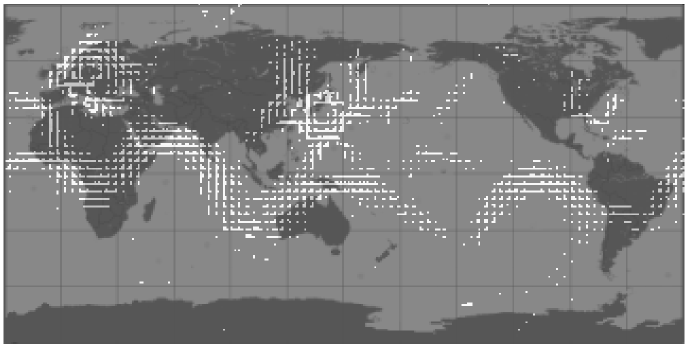
In April 1994, there are two densely distributed carbon dioxide concentration. One is situated in European countries and the other one is situated in south east area. In these cases, LL component is situated at top left corner while HH is situated at the bottom right corner. Mean while LH and HL are situated at the top right corner and the bottom left corner, respectively. It is comprehensive that spatial variability of carbon dioxide distribution is estimated with the different level of the MRA of resultant images.

Using the resultant images, reconstruction process is applied to the 3D (horizontal, vertical, and time dimensions) carbon dioxide distributions without low frequency components which should represent relatively large changes. Namely, spatial frequency components of the spatial variability can be recognized with the reconstructed images. The reconstructed images of each level are shown in Fig. 6 (a), (b), (c), and (d). In the reconstructed images, spatially calm and steeply changed areas can be detected depending upon the degree of changes.

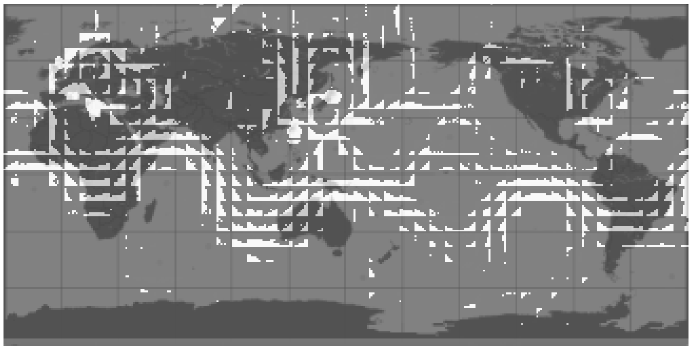
Meanwhile, carbon dioxide changes in time is tried to extract from the 3D concentration of the monthly average of carbon dioxide distribution images for 11 years which is shown in Fig. 7 with the different types of base function of MRA. Daubechies base function is superior to Haar base function for these experiments in terms of detailed information on large changes of carbon dioxide concentration distribution in time.



(a)Level1



(b)Level2



(C)Level3

Figure 6 Time changes detection with the different levels of MRA.

For 11 years, there are some significantly changed areas. These are not only European countries, but also middle and southern portion of African continents and Middle Eastern countries areas. There is the stream shape of large changed areas which is situated from El Nino offshore area and it goes to middle southern portion of South American continent as well as it goes through Atlantic ocean to the middle southern portion of African continent and it goes up to the middle eastern countries, after that it goes down to the western portion of Australian continent through Indian ocean and it goes up to Japanese island and the Pacific ocean area as well as it is divided into the other way to the east-ward and then goes down to the middle northern portion of the Pacific ocean are found.

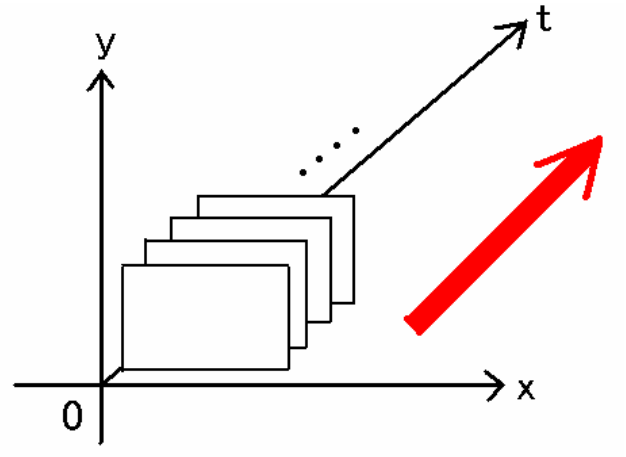


Figure 7 Time series of spatial distribution of carbon dioxide concentration distributions.

Moreover, large changed area is found at east coastal area of Asian continent. Furthermore, the detected these large changed areas are recognizable depending on the change frequency component which corresponds to the levels of MRA.

The reconstructed image without LLL component cannot be shown in figures, so that one dimensional reconstructed trend of carbon dioxide changes is illustrated in Fig. 8. The reconstructed data at the location of 13 degree south, 46 degree west (Tanzania eastern offshore) shows relatively large changes in time at the around 70 and 100 months after January 1990 which are corresponding to 1993 and 1997 of the El Nino years. In this case, level 1 of MRA with Daubechies base function is applied to the original carbon dioxide concentration distribution of image. Also Haar of base function is attempted to the original image. The results from the Daubechies base function are superior to that of Haar base function. The details of the large changed areas detected with Daubechies base function is much clear rather than that of Haar base function.

IV. Conclusions

Contiguous dataset of carbon dioxide concentration distribution can be derived from the sparsely situated data in space domain by using distant dependent linear interpolation.

Through an experiment with 11 years of carbon dioxide concentration data starting from 1990 provided by WDCGG: World Data Center for Greenhouse Gasses, it is found that there exists seasonal change and relatively large changes are occurred in El Nino years. It is also found that the carbon dioxide is concentrated in European continent.

It is found that the proposed 3D wavelet based MRA is useful to detect large changes of carbon dioxide concentration distribution in space and time domains. Also Daubechies base function is superior to Haar base function for details of the detected large changes. Level of MRA depends on the frequency components of large changes which would like to be detected.

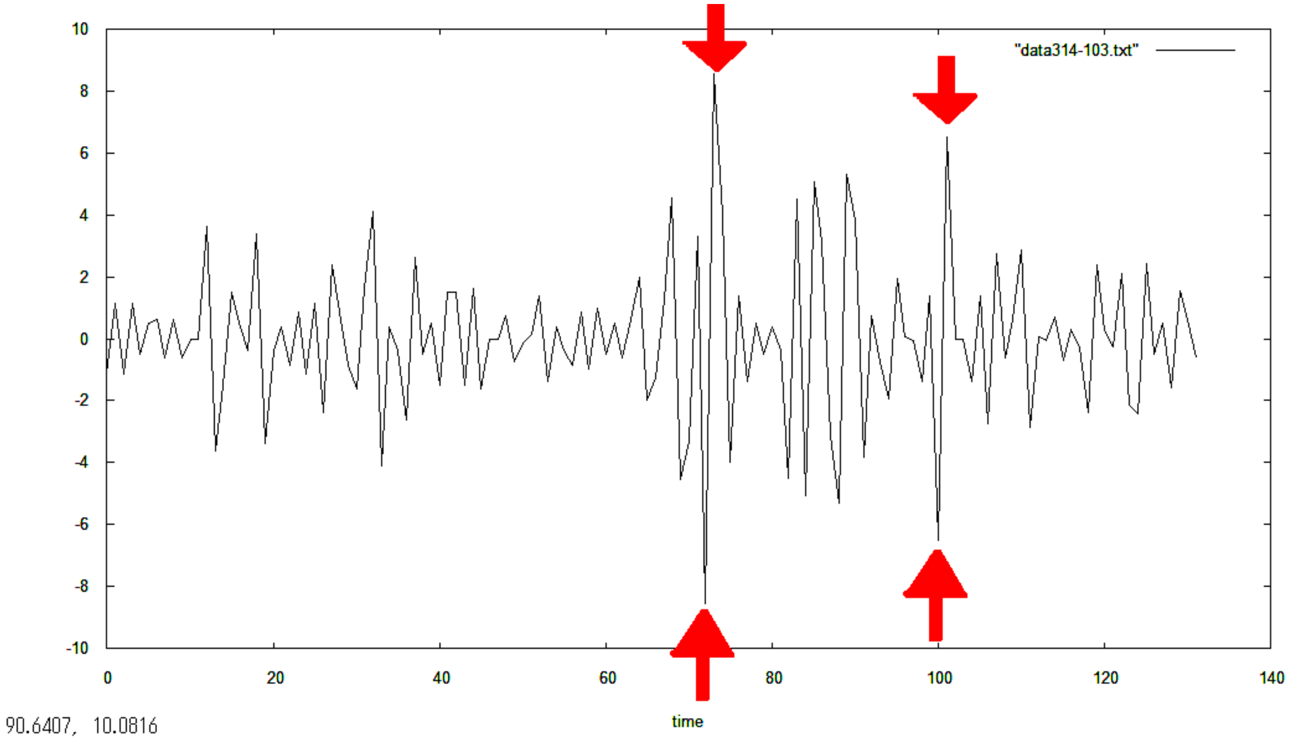


Figure 8 Time changes detected with wavelet based MRA with Daubechies base function and with level 3 which is applied to the original one dimensional time series of carbon dioxide concentration distribution data at the eastern portion of Tanzania. El Nino phenomena are observed at not only El Nino offshore of Peru but also the other stream lines of areas including the eastern portion of Tanzania.

ACKNOWLEDGMENT

The author would like to thank to Yoshikazu Saitoh for his experimental efforts.

REFERENCES

[1] C. L. Sabine et al., "The Oceanic Sink for Anthropogenic CO2" , Science 305, 367 (2004).

[2] R. F. Keeling, Comment on "The Ocean Sink for Anthropogenic CO2" Science 308, 1743 (2005);

www.sciencemag.org/cgi/content/full/308/5729/1743c

[3] Gurney K.R., R.M. Law, A.S.Denning, P.J.Rayner et.al, Towards robust regional estimates of CO2 sources and sinks using atmospheric transport models, Nature, 415, p626-630, 2002.

[4] K. Arai, Fundamental theory for wavelet analysis, Morikita Shuppan Publishing Co., Ltd., 2000.

[5] http://gaw.kishou.go.jp/wdcgg.html

[6] The World Data Centre for Greenhouse Gases (WDCGG) is established under the Global Atmosphere Watch (GAW) programme to collect, archive and provide data for greenhouse (CO_2, CH_4, CFCs, N_2O, etc.) and related (CO, NO_X, SO_2, VOC, etc.) gases and surface ozone in the atmosphere and ocean, measured under GAW and other programmes

AUTHORS PROFILE

Kohei ARAI

Saga University

Saga, Japan

Kohei Arai received a PhD from Nihon University in 1982. He was subsequently appointed to the University of Tokyo, CCRS, and the Japan Aerospace Exploration Agency. He was appointed professor at Saga University in 1990. He is also an adjunct professor at the University of Arizona and is Vice Chairman of ICSU/COSPAR Commission A.

Bidirectional WDM-Radio over Fiber System with Sub Carrier Multiplexing Using a Reflective SOA and Cyclic AWGs

Fady I. El-Nahal
Electrical Engineering Department, Faculty of Engineering
Islamic University of Gaza
Gaza City, Gaza Strip

Abstract— **A bidirectional SCM-WDM RoF network using a reflective semiconductor optical amplifier (RSOA) and cyclic arrayed waveguide gratings (AWGs) was proposed and demonstrated. The purposed RoF network utilizes Sub Carrier Multiplexed (SCM) signals for down-link and an on-off keying (OOK) signals re-modulated for up-link. In this paper, A 50 km range colorless WDM-ROF was demonstrated for both 1 Gbit/s downstream and upstream signals. The BER performance of our scheme shows that our scheme is a practical solution to meet the data rate and cost-efficient of the optical links simultaneously in tomorrow's ROF access networks.**

Keywords- Radio over Fiber (RoF); wavelength-division multiplexing (WDM); Sub-carrier modulation (SCM); arrayed waveguide grating (AWG).

I. INTRODUCTION

Recently, data capacity of wireless communication has been radically expanded from voices and simple messages to multimedia in order to satisfy various demands of system users with evolutionary future services. RoF systems could be the answer to many urgent needs of the telecommunication networks, as they could provide the necessary bandwidth for the transmission of broadband data to end-users, other benefits are low attenuation loss, and immunity to radio frequency interference [1-4]. The combination of sub-carrier multiplexing (SCM) and wavelength division multiplexing (WDM) is likely to play great role in these systems [5-8]. Wavelength division multiplexing (WDM) technique is used to simplify the network architecture by allowing different BSs to be fed using a single fiber. Each BS would be assigned its own wavelength and signals sent to different users serviced by a particular BS would be transmitted on that wavelength by means of SCM [9]. In a RoF system, Most of the signal processing processes (including coding, Multiplexing, and RF generation and modulation) are carried out By Central Office (CO), which makes the Base Station (BS) cost-effective. Therefore, RoF will become a key technology in the next generation mobile communication system [10-16].

For the uplink, the BSs must include an optical source, which is modulated by the mm-wave uplink radio signals. This approach results a high cost BS. Here we use a low cost uplink configuration, which eliminates the need for an expensive WDM source at the BS. This is accomplished by using a reflective semiconductor optical amplifier (RSOA) in the BS which replaces the high cost WDM source [17,18]. Many bidirectional RoF systems have been studied recently where a reflective semiconductor optical amplifier (RSOA) plays an important role. The RSOA can be used as a modulator and amplifier [9]. This approach avoids the need of stabilized a laser at each BS, this uplink configuration may be applied to other wireless networks such as 3G mobile communication systems. This system has the advantage of a simplified BS design but it brings the problem of dispersion caused fading that occurs when high frequency signals travel along fiber. Chromatic dispersion severely limits the transmission distance. Optical single sideband (OSSB) modulation techniques are used to overcome fiber dispersion effects [2,9,14].

In this paper, we propose a new self-restorable architecture for bidirectional WDM-PON. It utilizes one different wavelength assignments and 1 × N AWGs, as 16 channels each channel of AWGs is coupled to Base stations (BSs) by using a reflective semiconductor optical amplifier (RSOA). This is for achieving sixteen 1-Gb/s downstream and upstream signals.

II. WDM-ROF ARCHITECTURE

Figure 1 shows the proposed WDM-RoF architecture for transmitting subcarrier multiplexing (SCM) encoded channels over a bidirectional single mode optical fiber (50-km). At the central office (CO), a series of narrow bandwidth continuous wave (CW) with various wavelengths are modulated via a LiNbO3 Mach-Zehnder modulator using 1 Gb/s non-return to zero (NRZ) downstream data to generate downstream signals. Downlink data signal is mixed with local oscillator signal (10-GHz) and a carrier generator having a number of RF subcarriers. The generated SCM signals are multiplexed by Arrayed Waveguide Grating (AWG) and sent over the bidirectional single-mode fiber (SMF). A circulator is used in the central office (CO) to separate the downstream and upstream traffic. The SCM signals are de-multiplexed by AWG in remote node (RN) where various wavelength lights are sent to different Base stations. Simple AWGs that support both dedicated wavelengths and power-splitting bandwidth sharing are used at the CO and the remote node (RN).

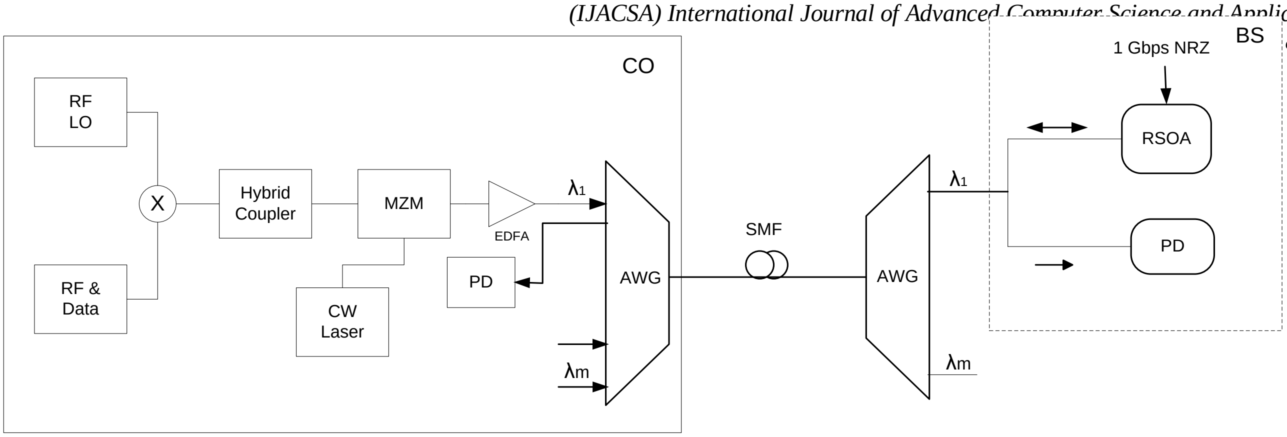


Figure 1: the proposed WDM-RoF architecture

At the BS, using optical splitter/coupler, portion of the SCM signal is fed to a SCM receiver. For up-link, the other portion of the downstream SCM signal from the splitter/coupler is re-modulated using 1 Gb/s NRZ upstream data by RSOA in the BS. The re-modulated OOK signals are sent back over the fiber to the CO where they are de-multiplexed by an AWG DEMUX. The reflected optical signal is detected by a PIN-photodiode. Uplink optical sidebands produce crosstalk when uplink data was detected at central station. Crosstalk can be reduced by using Bessel filter.

III. RESULTS AND DISSCUSSIONS

The WDM-RoF architecture was modeled using a commercially available package [19]. The proposed scheme uses SCM signal for downstream and OOK signal re-modulated by the RSOA for upstream. The received eye diagrams of downstream and upstream signals were measured at Base station and central office respectively. The received eye diagrams of the downlink and uplink signals are shown in Figure 2 and Figure 3 respectively. The results show that the Eye closure penalty is smaller for the uplink than that of the downlink which is expected, as the signal travel twice the distance for the uplink. Chromatic dispersion induced by bidirectional fiber will not cause downlink microwave signal a power penalty problem. So, the Maximum eye amplitude for downlink stage after signal transmission took place over 50-km of bidirectional fiber at base station.

BER simulations were carried out for both uplink and downlink with a Bit Rate of 1-Gb/s and no. of subcarriers = 70. The BER variation with input optical power Pin curves for the downlink and uplink are shown in figure 4. It is clear that both uplink and downlink do provide good BER performances, however the BER results for the downlink are better than those of the uplink. For example, when P_{in} = -5 dBm, the BER $=1.1\times10^{-11}$ for the downlink while it is 2.9×10^{-10} for the uplink. When P_{in} = +5 dBm, the BER $=8.8\times10^{-17}$ for the downlink while it is 1.1×10^{-12}

for the uplink. This can be attributed to the mixing noise between unsuppressed SCM signal in downstream and the digital signal of upstream which is generated in the remodulation process. This noise which influences the upstream signal could be reduced by using low pass filter after the photo detector in the CO. It is also noted from the figure that the BER for the uplink stays nearly constant for $P_{in} \geq -1$ dBm. This can be explained by the fact that the RSOA is operating in the gain saturation region. The variation of the gain of RSOA with the optical input power Pin is shown in figure 5. It is clear that the maximum gain appears at P_{in}= -5 dBm, then goes down to reach the lowest gain at P_{in}= -1 dBm where it goes into saturation.

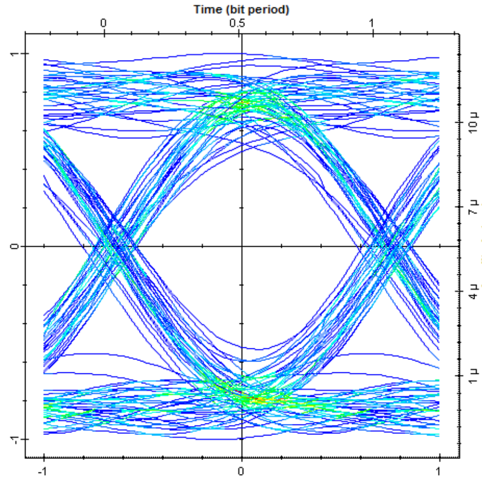


Figure 2: Eye diagram of downlink

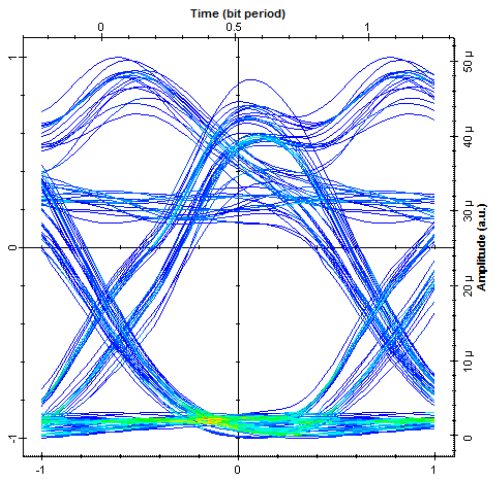


Figure 3: Eye diagram of uplink

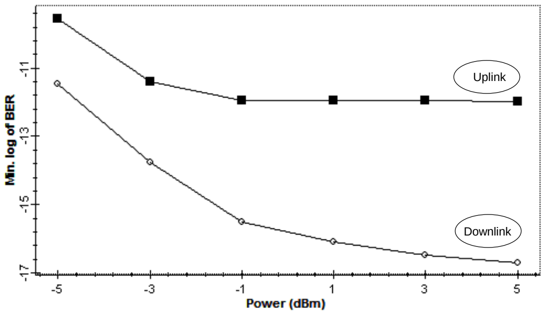


Figure 4: BER versus input power

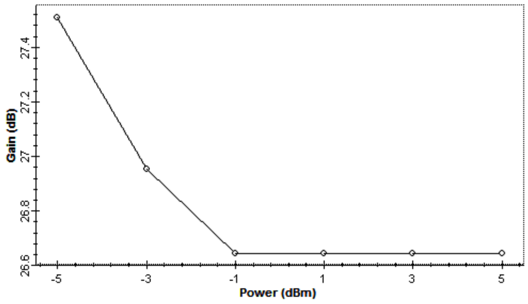


Figure 5: The variation of RSOA gain with input power.

The variation of BER with no. of subcarriers is plotted in Fig. 6. It is clear from the results that increasing the number of subcarrier channels degrade the system performance. For example, when RF channels increase from 10 to 90, corresponding BER decreases from 4.6×10^{-21} to 5.3×10^{-14} for the downlink, while it decreases from 1.4×10^{-15} to 7.3×10^{-13} for the uplink. This can be explained by the fact that, as the number of sub-carrier channels increases, their frequency spacing decreases which results crosstalk or noise.

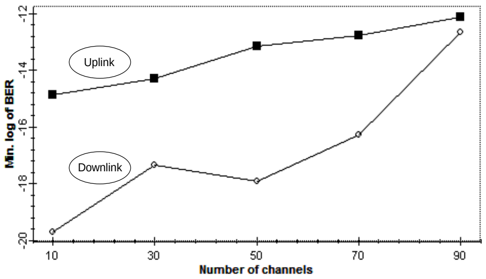


Figure 6: BER versus no. of channels

IV. CONCLUSION

The WDM-RoF model has been proposed as solution for increased bandwidth demand. The combination of WDM and SCM has been performed to provide high data rates and bandwidth in wireless communication. In this paper we have analyzed the performance of WDM/SCM Radio over Fiber System. We presented a demonstration of 1Gb/s signal for up/downstream in 50-km bidirectional link. The upstream traffic is obtained by remodulating the downstream traffic at the BS. The results obtained here show that increasing total number of sub-carriers channels has a significant impact on performance of WDM-SCM ROF system. The results show that WDM/SCM system can play great role in future RoF systems.

REFERENCES

[1] H. Nasoha and S. M. Idrus" Modeling and Performance Analysis of WCDMA Radio over Fiber System Applied Electromagnetic", APACE 2007. Asia-Pacific Conference, December 2007 Melaka.

[2] Hyun-Seung Kim et al. , "Bidirectional WDM-RoF Transmission for Wired and Wireless Signals", Proc. of SPIE-OSA-IEEE Asia Communications and Photonics, SPIE Vol. 7632, 76322F, (2009).

[3] A Nirmalathas, P A. Gamage, C Lim, D Novak, and R Waterhouse, “Digitized Radio-Over-Fiber Technologies for Converged Optical Wireless Access Network”, Journal of Lightwave Technology, Vol. 28, No. 16, pp. 2366-2375, (2010).

[4] Y.-M. Lin and P.-L. Tien, “Next-Generation OFDMA-Based Passive Optical Network Architecture Supporting Radio-Over-Fiber”, IEEE Journal on Selected Areas in Communications - JSAC , vol. 28, no. 6, pp. 791-799, (2010).

[5] A. Kaszubowska, L.P. Barry, P. Anandarajah and Ling Hu "Characterization of Wavelength Interleaving in Radio-over-Fiber Systems Employing WDM/SCM", Optics Communications, 260 (1). pp. 144-149. ISSN 0030-4018, 2006.

[6] O.K. Tonguz et al., J. Lightwave Technol. 14 (1996) 1400.

[7] A. Stohr et al., Electron. Lett. 35 (1999) 1653.

[8] Fulvio Grassi, José Mora, Beatriz Ortega, José Capmany, “Experimental Evaluation of the Transmission in a Low Cost SCM/WDM Radio over Fibre System Employing Optical Broadband Sources and Interferometric Structures”, in Proc. ICTON-09, pp. 1-4, São Miguel Island.

[9] M.C.R. Medeiros, R. Avó, P. Laurêncio, N.S. Correia, A. Barradas, H.J.A. da Silva, I. Darwazeh, J.E. Mitchell and P.M.N. Monteiro. “Radio

[10] over Fiber Access Network Architecture Employing Reflective Semiconductor Optical Amplifiers", in Proc. ICTON-MW'07, pp. 1-5, Sousse, Tunisia.

[11] L. RAO, X. SUN, W. LI, D. HUANG "OFDM-ROF System and Performance Analysis of Signal Transmission" Optoelectronics, 2006 Optics Valley of China International Symposium on, Nov., 2006.

[12] R. Hui, Benyuan Z, Renxiang H., C. T. Allen, Kenneth R.D, and Douglas R. "Subcarrier Multiplexing for High-Speed Optical Transmission", Journal of Lightwave Technology, VOL. 20, NO. 3, MARCH, 2002.

[13] C. Loyez, C. Lethien, R. Kassi, J.P. Vilcot, D. Decoster, N. Rolland and P.A. Rolland "Subcarrier radio signal transmission over multimode fibre for 60 GHz WLAN using a phase noise cancellation technique" ELECTRONICS LETTERS 20th Vol. 41 No. 2 Jan., 2005

[14] E. J. Tyler, P. Kourtessis, M. Webster, E. Rochart, T. Quinlan, S. E. M. Dudley, S. D. Walker, R. V. Penty, and I. H. White, "Toward Terabit-per-Second Capacities Over Multimode Fiber Links Using SCM/WDM Techniques ", Journal of Lightwave Technology, vol. 21, no. 12, Dec., 2003.

[15] A. Kaszubowska, P. Anandarajah, and L. P. Barry "Multifunctional Operation of a Fiber Bragg Grating in a WDM/SCM Radio Over Fiber Distribution System" IEEE Photonics Technology Letters, vol. 16, no. 2, Feb., 2004.

[16] S. M. Idrus,. "Photoparametric Amplifier in microwave Subcarrier Multiplexed Systems (PPA -in-SCM)", 22 Aug.,t 2007.

[17] T. Kim, K. Jeung-Mo, and H. Sang-Kook "performance analysis of bidirectional hybrid WDM/SCM PON link based on reflective semiconductor optical amplifier" Microwave and optical technology letters , ISSN 0895-2477, Vol. 48, April, 2006.

[18] Zhansheng Liu et al., "Experimental Validation of a Reflective Semiconductor Optical Amplifier Model Used as a Modulator in Radio Over Fiber Systems", IEEE Photon. Technol. Lett., vol. 23, no. 9, pp. 576–578, (2011).

[19] Yong-Yuk Won et al. , "Full Colorless WDM-Radio Over Fiber Access Network Supporting Simultaneous Transmission of Millimeter-Wave Band and Baseband Gigabit Signals by Sideband Routing", Journal of Lightwave Technology, vol. 28, no. 16, pp. 2213–2218, (2010).

[20] Optisystem from optiwave.

AUTHORS PROFILE

Dr. Fady I. El-Nahal received his B.Sc. degree in electrical and electronic engineering in 1996 from Alfateh University and his M.Phil. and Ph.D. degrees from the University of Cambridge in 2000 and 2004 respectively. He is currently with the Department of Electrical Engineering, The Islamic University of Gaza. His research activities include optoelectronics, optical communications and wavelength routing in optical networks. Fady is a Fellow of the Cambridge Overseas Trust and the chairman of the Oxford and Cambridge Society of Palestine.

AODV Robust (AODV$_R$): An Analytic Approach to Shield Ad-hoc Networks from Black Holes

Mohammad Abu Obaida[1]
Dept. of CSE
DUET
Gazipur, Bangladesh

Shahnewaz Ahmed Faisal[2]
Dept. of CSE
KUET
Khulna, Bangladesh

Md. Abu Horaira[3]
Software Engineer
DataSoft BD Ltd
Chittagong, Bangladesh

Tanay Kumar Roy[4]
Dept. of CSE
KUET
Khulna, Bangladesh

***Abstract*—Mobile ad-hoc networks are vulnerable to several types of malicious routing attacks, black hole is one of those, where a malicious node advertise to have the shortest path to all other nodes in the network by the means of sending fake routing reply. As a result the destinations are deprived of desired information. In this paper, we propose a method AODV Robust (AODV$_R$) a revision to the AODV routing protocol, in which black hole is perceived as soon as they emerged and other nodes are alerted to prevent the network of such malicious threats thereby isolating the black hole. In AODV$_R$ method, the routers formulate the range of acceptable sequence numbers and define a threshold. If a node exceeds the threshold several times then it is black listed thereby increasing the network robustness.**

Keywords- Ad-hoc Networks; Wireless Networks; MANET; RT; AODV; Ad-hoc Optimal Distance Vector; Black-hole; OPNET.

I. INTRODUCTION

Ad-hoc networks are exemplified by dynamic topology, self-configuration, self-organization, constrained power, transitory network and lack of infrastructure. Characteristics of these networks lead to using them in disaster recovery operation, smart buildings and military battlefields [3].

Mobile Ad-hoc Network (MANET) routing protocols are classified into two basic classes, proactive and reactive [2]. In proactive routing protocols, routing information of nodes is exchanged intermittently, such as DSDV [4]. However, in on-demand routing protocols nodes exchange routing information as required such as, AODV [1] and DSR [5]. The AODV routing protocol [13] is an adaptation of the DSDV protocol for dynamic link conditions.

AODV is used to find a route between source and destination as required and this routing protocol uses three significant type of messages, route request (RREQ), route reply (RREP) and route error (RERR). Ground information of these messages, such as source sequence number, destination sequence number, hop count and etc is explicated in feature in [1]. Each of the nodes has a routing table (RT), which contains information about the route to the particular destination. When source node desires to communicate with the destination and if in routing table there is no route between, source node broadcasts RREQ initially. As RREQ is received by intermediate nodes that are in the transmission range of sender, those nodes broadcast RREQ until RREQ is received by destination or an intermediate node that has fresh enough route to the destination. Then it sends RREP unicastly toward the source. As a result, a route between source and destination is established. A fresh enough route is a valid route entry that its destination sequence number is at least as great as destination sequence number in RREQ. The source sequence number is used to determine freshness about route to the source consequently destination sequence number is used to determine freshness of a route to the destination. When intermediate nodes receive RREQ, with consideration of source sequence number and hop count, make or update a reverse route entry in its routing table for that source. Furthermore, when intermediate nodes receive RREP, with consideration of destination sequence number and hop count, make or update a forward route entry in its routing table for that destination.

Though reliable environments have been assumed in the majority of researches on ad-hoc routing protocols, unreliable situations are quite often. Therefore, most ad-hoc routing protocols are susceptible to miscellaneous types of attacks such as Spoofing attack, Denial of Service (DoS) attack, Routing Loop attack, Warm hole attack [6], Black hole attack etc. Common types of threats are possessed against Physical, MAC and Network layer, that are the fundamental layers requires for proper functioning of routing protocol. The threats try to accomplish two purposes: not forwarding the packets or add/alter some parameters (e.g. sequence number or hop count) to routing messages. In Black hole attack, a malicious node uses the routing protocol to advertise itself as having the shortest or freshest path to the node whose packets it wants to intercept. In a flooding based protocol, the attacker eavesdrops to requests for routes. When the attacker receives a request for a route to the target node, it creates a reply consisting of an exceptionally short or fresh route [7], therefore, misleading the source in transferring information to the path that leads to the black hole itself.

Intrusion detection is a challenging task in MANETs. Zhang and Lee [8] propose a circulated and cooperative intrusion detection model based on statistical incongruity detection techniques. Dang et. al. [9] introduces a method that requires each of the intermediate nodes to send back the next hop information inside RREP message. This method uses further request message and further reply message to confirm the authority of the route. In Robust Routing [10] by Lee, Han, Shin, the intermediate node requests its next hop to send a confirmation message to the source. After receiving both route reply and authentication message, the source verifies the legitimacy of path according to its policy. An approach based on dynamic training method in which the training data is updated at regular time intervals has been proposed Kurosawa et. al. in [11]. In [12], Huang et al use both specification-based and statistical-based approaches. They construct an Extended Finite State Automation (EFSA) according to the specification of AODV and model normal state and detect attacks with incongruity detection and specification-based detection.

With the view to secure routing in MANET several intelligible researches has been carried out. Hu, and Johnson proposed SEAD [14], a secure routing protocol based on DSDV that employs Hash chains to authenticate hop counts and sequence numbers. ARAN [15] harnesses cryptographic public-key certificates in order to accomplish the security target. A modified Ad-hoc routing protocol has been proposed by Ariadne [16] that provides security in MANET and depends on efficient symmetric cryptography. Secure AODV (SAODV) [17] is a security extension of AODV protocol, based on public key cryptography. Hash chains are used in this protocol to authenticate the hop count. Adaptive SAODV (A-SAODV) [18] has proposed a mechanism based on SAODV for improving the performance of SAODV. In [19] a bit of modification has been applied to A-SAODV for increasing its performance.

II. Black Holes: A Network Layer Attack in MANET

In black hole attack, the malicious node waits for the neighbors to initiate a RREQ. Obtaining the RREQ right away it sends a false RREP with a modified higher sequence number. As a result, the source node assumes that node (malicious) is having the fresh route towards the destination.

The source node ignores the RREP packet received from other nodes and begins to send the data packets over malicious node. In this way, the black hole swallows all objects and data packets [20].

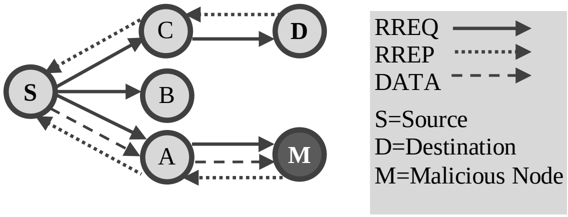


Figure 1. Black hole attack in a mobile ad-hoc network.

As demonstrated in figure 1, source node S requests to send data packets to destination D, Malicious Node M acts as a black hole replying with false reply RREP having higher modified sequence number. Accordingly, data communication initiates from S towards M instead of D.

Black hole attack in AODV protocol can be classified into two categories:

1) Black hole attack caused by RREQ

With sending fake RREQ messages an attacker can form black hole attack as follows:

a) Set the originator IP address in RREQ to the originating node’s IP address.

b) Set the destination IP address in RREQ to the destination node’s IP address.

c) Set the source IP address of IP header to its own IP address.

d) Set the destination IP address of IP header to broadcast address.

e) Choose high sequence number and low hop count and put them in related fields in RREQ.

So, false information about source node is inserted to the routing table of nodes that get sham RREQ. Hence, if these nodes want to send data to the source, at first step they send it to the malicious node.

2) Black hole attack caused by RREP

With sending fake RREP messages an attacker can form black hole attack. After receiving RREQ from source node, a malicious node can generate black hole attack by sending RREP as follow:

a) Set the originator IP address in RREP to the originating node’s IP address.

b) Set the destination IP address in RREP to the destination node’s IP address.

c) Set the source IP address of IP header to its own IP address.

d) Set the destination IP address of IP header to the IP address of node that RREQ has been received from it.

III. AODV$_R$: Approach against Black Hole Attacks

In AODV the node that receives the RREP, checks the value of sequence number in routing table and accepts if it has a higher RREP *seq_no* than the one in routing table.

```
IF (RREP seq_no > RT_seq_no) THEN
     RREP is ACCEPTED
ELSE
     RREP is DISCARDED
```

To solve this, we added an extra method to check whether the RREP seq_no is higher than the threshold value (A value that is updated dynamically in time intervals). As the value of RREP seq_no is found to be higher than the threshold value, the node is suspected to be malicious and added to the black list.

```
IF (RREP seq_no > THRESOLD) THEN
    Send ALARM to neighbors
ELSE
    RREP is ACCEPTED
```

The threshold value is dynamically updated using the data collected in the time interval. If the initial training data were used it is implausible for the routers to adapt changes in environment. The threshold value is the average of the difference of dest_seq_no in each time slot between the sequence number in the routing table and the RREP. If a node receives a RREP for the first time, it updates value of the threshold.

If max chances of aberration (RREP seq_no>THRESOLD) is detected, it sends a new control packet ALARM to its neighbors. The ALARM packet contains the black list node as a parameter that tells the neighboring nodes to discard RREP from that malicious node. Further if any node receives the RREP, it looks over the list to check if the reply is from the blacklisted node and simply ignores the node throughout communication if identified as black hole. In this way, the malicious node is isolated from the network that results in less routing overhead under threats. Moreover the design not only detects the black hole attack, but also prevents it further by updating threshold which reflects the real environment.

A. *Route Analyzer*

Route analyzer a module in router assumed to store the past routing history, i.e. the list of destination sequence number, hop count in each time slot. We find the average of increments in destination_sequence_no for the available time slots/ history, i.e. if dest_seq_no is assumed as an array; we find the difference in every pair of successive terms and average that values. This leaves us with a value that further is used to as minimum of threshold range.

Another arithmetic mean is considered that is the average between RREP_seq_no and RT_seq_no in each time frame (i) for destination. It is added with the previous min_threshhold value to find the maximum of the range.

$$\sum(RREP\_seq\_no_i - RT\_seq\_no_i\,) \,/\, Total\ no.\ of\ frames$$

It would not be fair to list a node as black for single aberration in provided destination sequence number or hop count. Such an action may lead the network to bareness because the topology is dynamic in Ad-hoc Networks. Instead we count the number of anomalies detected for any node. In addition, if the total number of deception detected reaches the aberration tolerance value than it is identified as black hole and neighbors are ALARMed.

B. *AODV$_R$ Process Development*

The proposed architecture AODV$_R$ demonstrated in the Figure 2 formed of several modules that are Packet Classifier, Extractor, Blacklist Tester, RREP sequence number Tester, Threshold Tester and ALARM broadcaster. As the packet arrives in the system Packet Classifier classifies it to be RREQ, RREPsecure, RERR, ALARM and HELLO packet. AODV$_R$ assumes format of RREQ, RERR and HELLO Packets are as same as the AODV. However it modifies the content and format of RREP and includes a new type of packet ALARM.

Extractor extracts required contents of all types of packets other than HELLO. Three diamonds including threshold tester as depicted in the process flow of figure 2 check whether the packets are from a reliable source or not and discards the node or packet accordingly.

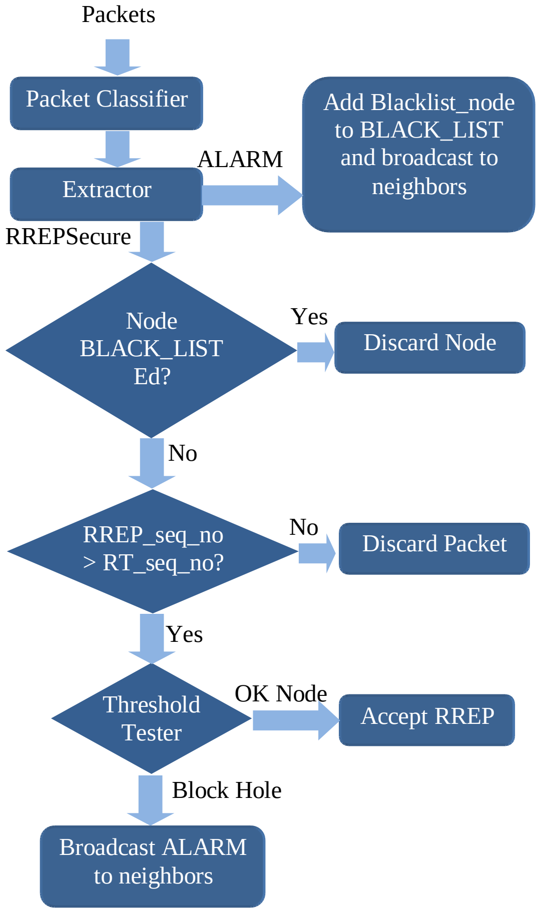


Figure 2. AODV$_R$ Process Development.

Every of the nodes are given MAX_ABBERATION_TOLERANCE number of chances before they are attributed as BLACK_LISTed node; if an aberration is noticed than the node is check over and over before it emulates maximum chances. As a node is identified as black hole, ALARM Broadcaster broadcasts alert to neighboring nodes with the BLACK_LIST node as parameter. Any router receiving the ALARM packet forwards the message to its neighboring nodes thereby discovering the BLACK_LIST to the whole network.

C. *RREPsecure & ALARM*

```
Start
  Packet classifier←PACKET

  IF (PACKET=RREPsecure) THEN
     RREP_seq_no := Packet extractor←RREPsecure
  PACKET
     Blcklist←Check if the source_addr is in BLACK_LIST

     IF (Node is BLACK_LISTed) THEN
         Simply IGNORE the node.
     ELSE
         IF (RREP_seq_no > RT_seq_no) THEN
```

```
            IF (RREP_seq_no > THRESOLD_VALUE)
            THEN
               IF (NODE_CHANCES <
               MAX_ABBERATION_TOLERANCE)
               THEN
                  NODE_CHANCES :=
                  NODE_CHANCES+1;
                  Recheck the authenticity of the node by
                  RREQ.
               ELSE
                  Broadcast ALARM to neighbors
            ELSE
                  ACCEPT RREP and FORWARD
         ELSE
            DISCARD RREP

   ELSE IF (PACKET=ALARM) THEN
      Blacklist_node := Packet extractor←ALARM packet
      Add Blacklist_node with BLACK_LIST
      Broadcast ALARM to neighbors
    Stop.
```

IV. Performance Evolution

We implemented AODV$_R$ in OPNET [21] simulator and evaluated the performance based on three parameters that are Packet Delivery Ratio (PDR), Average End-to-End Delay (Avg E-E Delay) and Normalized Routing Overhead (NRO). PDR is the ratio of data delivered to the destination to data sent out by the source and Avg E-E Delay is the delay caused by the transmission.

We have considered various network contexts that were formed by varying Network Size, Traffic Load (total sources), and Mobility for the purpose of proper evolution.

A. Impact of Mobility

We evaluated the performance of AODV normal, AODV under attack and AODV$_R$ under attack in the context of variation in mobility that are listed in Table I (PDR) and Table II (Avg E-E Delay) and depicted consequently in Figure 3 and Figure4.

TABLE I. PDR (%) VS MOBILITY (m/s) FOR AODV & AODV$_R$

Method	Mobility(m/s)							
	10	20	30	40	50	60	70	80
AODV normal	100	98	92	86	85	84	86	89
AODV under attack	10	18	16	20	25	26	42	45
AODV$_R$ under attack	97	96	88	86	84	80	82	83

As illustrated in figure 3, AODV results in very low PDR under attack while AODV$_R$ exhibits almost same capability (3%-5% ranging from AODV) as normal AODV does. Later, Figure 4 testimonies AODV$_R$ to be delay efficient.

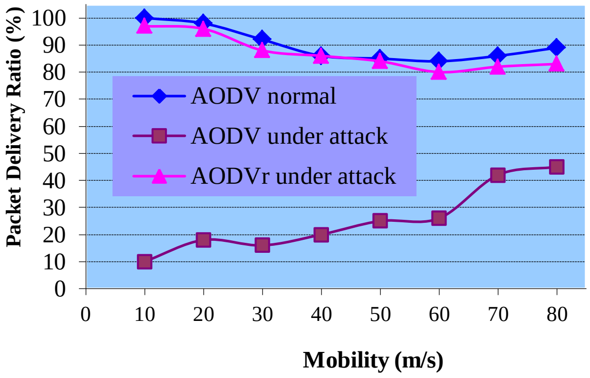


Figure 3. Graph of PDR (%) vs Mobility (m/s) for data in Table 1

TABLE II. AVERAGE END-TO-END DELAY VS MOBILITY (m/s)

Method	Mobility(m/s)							
	10	20	30	40	50	60	70	80
AODV normal	0.01	0.03	0.05	0.05	0.06	0.06	0.07	0.07
AODV$_R$ under attack	0.01	0.04	0.04	0.06	0.06	0.06	0.07	0.07

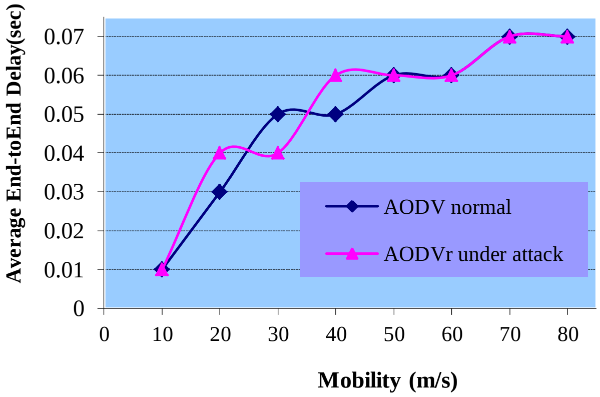


Figure 4. Graph of Average End-toEnd Delay vs Mobility (m/s)

B. Impact of Network Size

Performance of AODV normal, AODV under attack and AODV$_R$ under attack are evaluated in the circumstance of discrepancy in network size (no. of nodes) that are listed in Table III (PDR), Table IV (Avg E-E Delay), Table V (Normalized Routing Overhead) and delineated accordingly in Figure 5, Figure 6, and Figure 7.

TABLE III. PDR (%) VS NETWORK SIZE IN AODV & AODV$_R$

Method	Total nodes					
	10	20	30	40	50	60
AODV normal	100	99	96	97	98	99
AODV under attack	18	16	19	15	17	15
AODV$_R$ under attack	100	95	96	97	95	98

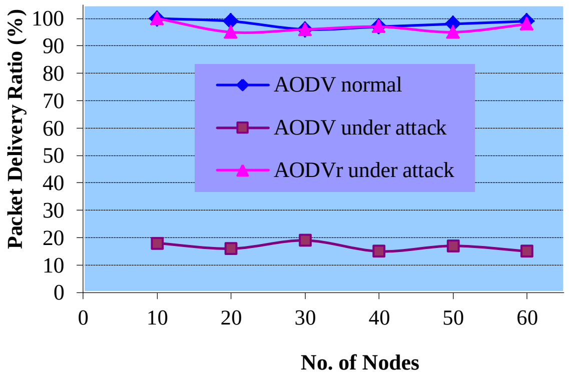


Figure 5. Graph of PDR (%) vs Network Size

TABLE IV. AVERAGE END-TO-END DELAY VS NETWORK SIZE

Method	Total no. of nodes					
	10	20	30	40	50	60
AODV Normal	0.017	0.046	0.048	0.05	0.052	0.052
AODV$_R$ under attack	0.018	0.043	0.046	0.048	0.051	0.052

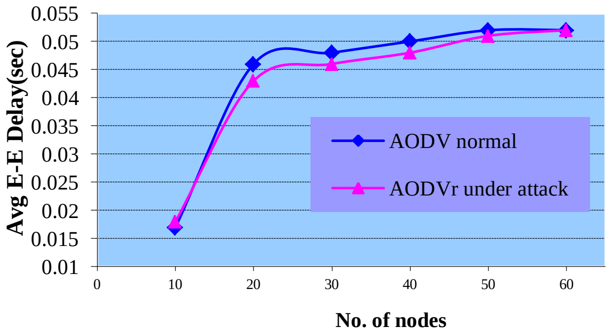


Figure 6. Graph of Average End-to-End Delay Vs Network Size

TABLE V. NORMALIZED ROUTING OVERHEAD (NRO) VS NETWORK SIZE

Method	Total no. of nodes					
	10	20	30	40	50	60
AODV normal	0.01	0.08	0.19	0.23	0.24	0.28
AODV$_R$ under attack	0.0	0.11	0.17	0.24	0.25	0.28

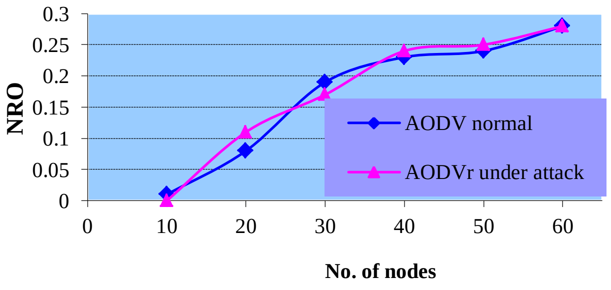


Figure 7. Graph of NRO Vs Network Size

In case of variation in Network size, as demonstrated in figure 5, AODV results in very low PDR under attack however AODV$_R$ exhibit s almost same performance as AODV does. Subsequently, Figure 6 manifests AODV$_R$ to be delay efficient however trivial falls that are negligible. Later Figure 7 testimonies a small increase in NRO that is insignificant.

C. Impact of Traffic Load

We simulated the performance of AODV, AODV under attack and AODV$_R$ under attack in the circumstance of discrepancy in Traffic Load (no. of sources) that are listed in Table VI (PDR), Table VII (Avg E-E Delay), Table VIII (NRO) and depicted accordingly in Figure 8, Figure 9, and Figure 10.

TABLE VI. PDR (%) VS TRAFFIC LOAD IN AODV & AODV$_R$

Method	No. of Sources					
	1	2	3	4	5	6
AODV normal	100	95	90	88	80	82
AODV under attack	10	35	30	30	30	31
AODV$_R$ under attack	100	92	90	80	80	81

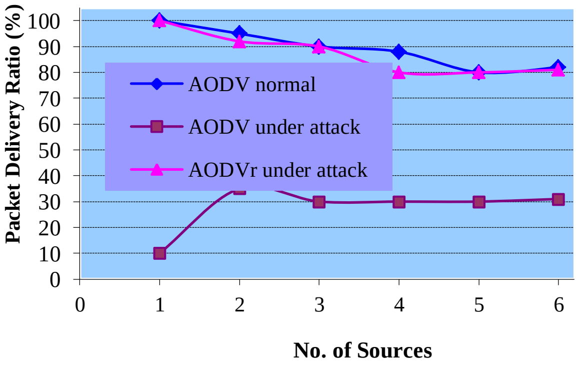


Figure 8. Graph of PDR(%) Vs Traffic Load

TABLE VII. AVG E-E DELAY VS TRAFFIC LOAD

Method	No. of Sources					
	1	2	3	4	5	6
AODV normal	0.04	0.08	0.12	0.14	0.22	0.25
AODV$_R$ under attack	0.05	0.08	0.12	0.13	0.20	0.23

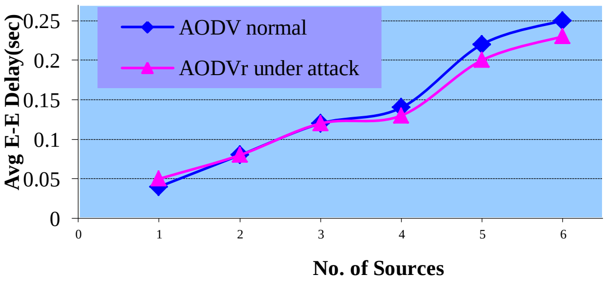


Figure 9. Graph of Average End-to-End Delay Vs Traffic Load

TABLE VIII. NRO Vs TRAFFIC LOAD

Method	**No. of Sources**					
	1	2	3	4	5	6
AODV normal	0.08	0.14	0.20	0.20	0.22	0.24
AODV$_R$ under attack	0.08	0.16	0.21	0.22	0.23	0.25

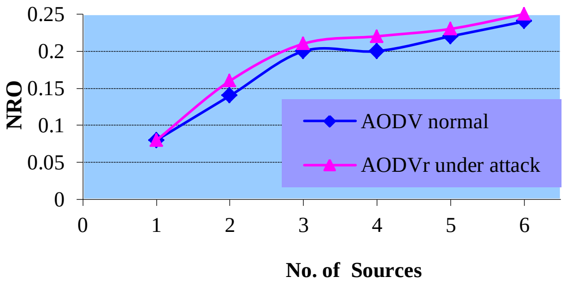


Figure 10. Graph of NRO Vs Traffic Load

In case of different Traffic Load, as depicted in figure 8, it is clear that as the traffic load increases the PDR of AODV$_R$ increases by 60% than AODV under attack that is very close to PDR of AODV normal. Afterward, Figure 9 shows AODV$_R$ to be delay efficient and sometimes better than AODV. Later on Figure 10 demonstrates a small NRO increment that can be ignored without hesitation.

V. CONCLUSION

Proposed AODV$_R$ exhibits appreciable performance dealing with networks with black holes; however the procedure of formulating the threshold is a bit overwhelming. Formulations of correct threshold range keep black holes from intrude; while a wrong formulation may restrict an authentic node thereby disgrace it to be a black hole.

Hence, this value has to be calculated and verified suitably.

REFERENCES

[1] C. E. Perkins, E. M. B. Royer and S. R. Das, "Ad-hoc On-Demand Distance Vector (AODV) Routing," Mobile Ad-hoc Networking Working Group, Internet Draft, draft-ietf-manetaodv- 00.txt, Feb. 2003.

[2] E. M. Royer and C-K Toh, "A Review of Current Routing Protocols for Ad-Hoc Mobile Wireless Networks," IEEE Person. Commun., Vol. 6, no. 2, Apr. 1999.

[3] E. Çayırcı, C.Rong, "Security in Wireless Ad Hoc and Sensor Networks," vol. I. New York: Wiley 2009, pp. 10.

[4] C. E. Perkins and P. Bhagwat, "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers," Proceedings of the SIGCOMM '94 Conference on Communications Architectures, Protocols and Applications, pp 234–244, Aug.1994.

[5] D. B. Johnson, D. A. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks," Mobile Computing, edited by Tomasz Imielinski and Hank Korth, Chapter 5, pp 153- 181, Kluwer Academic Publishers, 1996.

[6] Y. C. Hu, A. Perring, D. B. Johnson, "Wormhole Attacks in Wireless Networks," IEEE Journal On Selected Areas In Communications, Vol. 24, No. 2, Feb. 2006.

[7] M. Ilyas, "The Handbook of Ad hoc wireless Networks," CRC Press, 2003.

[8] Y. Zhang and W. Lee, "Intrusion detection in wireless ad-hoc networks," 6th annual international Mobile computing and networking Conference Proceedings, 2000.

[9] H. Deng, W. Li, and D. P. Agrawal, "Routing security in ad hoc networks," IEEE Communications Magazine, vol. 40, no. 10, pp. 70-75, Oct. 2002

[10] S. Lee, B. Han, and M. Shin, "Robust routing in wireless ad hoc networks," in ICPP Workshops, pp. 73, 2002.

[11] S. Kurosawa, H. Nakayama, N. Kat, A. Jamalipour, and Yoshiaki Nemoto, "Detecting Blackhole Attack on AODV-based Mobile Ad Hoc Networks by Dynamic Learning Method," International Journal of Network Security, Vol.5, No.3, pp 338-346, Nov. 2007.

[12] Y. A. Huang and W. Lee, "Attack analysis and detection for ad hoc routing protocols," in The 7th International Symposium on Recent Advances in Intrusion Detection (RAID'04), pp. 125-145, French Riviera, Sept. 2004.

[13] Tamilselvan, L.; and Sankaranarayanan, V. (2007). "Prevention of blackhole attack in MANET," The 2nd International Conference on Wireless Broadband and Ultra Wideband Communications. AusWireless, 21-21.

[14] Y.-C. Hu, D. B. Johnson, A. Perrig, "SEAD: Secure Efficient Distance Vector Routing for Mobile Wireless Ad hoc Networks," Proc. 4th IEEE Workshop on Mobile Computing Systems and Applications, Callicoon, NY, Jun. 2002, pp. 3-13.

[15] K. Sanzgiri, B. Dahill, B. Levine, C. Shields, and E. Belding-Royer, "A Secure Routing Protocol for Ad Hoc Networks," Proc. of IEEE International Conference on Network Protocols (ICNP), pp. 78-87, 2002.

[16] Y.-C. Hu, A. Perrig, D. B. Johnson, "Ariadne: A Secure On-Demand Routing Protocol for Ad hoc Networks," Proc. 8th ACM Int'l. Conf. Mobile Computing and Networking (Mobicom'02), Atlanta, Georgia, Sep. 2002, pp. 12-23.

[17] M. Zapata, "Secure Ad Hoc On-Demand Distance Vector (SAODV)," Internet draft, draft-guerrero-manet-saodv-01.txt, 2002.

[18] D. Cerri, A. Ghioni, "SecuringAODV: The A-SAODV Secure Routing Prototype," IEEE Communication Magazine, Feb. 2008, pp 120-125.

[19] K. Mishra, B. D. Sahoo, "A Modified Adaptive-Saodv Prototype For Performance Enhancement In Manet," International Journal Of Computer Applications In Engineering, Technology And Sciences (Ij-Ca-Ets), Apr. 2009 – Sep. 2009, pp 443-447.

[20] Chen Hongsong; Ji Zhenzhou; and Hu Mingzeng (2006). "A novel security agent scheme for AODV routing protocol based on thread state transition," Asian Journal of Information Technology, 5(1), 54-60.

[21] http://www.opnet.com

AUTHORS PROFILE

Mohammad Abu Obaida^1 obtained his Bachelor of Science in Engineering degree from Department of CSE, Dhaka University of Engineering & Technology (DUET), Gazipur-1700, Bangladesh. At present wide-ranging research on networks, cryptography and pattern recognition are carried out by him. His key research interests include Cryptography, Networks and Web Security, Wireless networks, Software Architecture, Machine Vision, Artificial Intelligence, Protocol analysis and design.

Control Systems application in Java based Enterprise and Cloud Environments – A Survey

Ravi Kumar Gullapalli
Hewlett-Packard
Bangalore, India

Dr. Chelliah Muthusamy
Yahoo
Bangalore, India

Dr.A.Vinaya Babu
JNTUH
Hyderabad, India

Abstract— **The classical feedback control systems has been a successful theory in many engineering applications like electrical power, process, and manufacturing industries. For more than a decade there is active research in exploring feedback control systems applications in computing and some of the results are applied to the commercial software products. There are good number of research review papers on this subject exist, giving high level overview, explaining specific applications like load balancing or CPU utilization power management in data centers. We observe that majority of the control system applications are in Web and Application Server environments. We attempt to discuss on how control systems is applied to Web and Application(JEE) Servers that are deployed in Enterprise and cloud environments. Our paper presents this review with a specific emphasis on Java based Web, Application and Enterprise Server Bus environments. We conclude with the future reserach in applying control systems to Enterprise and Cloud environments.**

Keywords- Control Systems; Java, Web Servers; Application Servers; Web Services; Enterprise Service Bus.

I. Introduction

Control Systems theory has been successfully applied to several engineering applications such as electrical power, process engineering, manufacturing plants. More than a decade there is an active research in investigating the applicability of control theory in different areas of the computing systems and data networks. It provides a systematic approach to achieve service level objectives by designing appropriate feedback control loops [1]. The control theory has been explored in various areas of computing such as web servers, multi-media systems, proxy caches, database servers, multi-tier web sites and real-time systems, power management of data centers [2]. The advantages of control theory are the self-managing or self-correction capabilities that make any physical or computing system to manage itself [3] [4]. This is achieved without continuous monitoring thereby avoiding huge manual intervention. Typically the software developments teams' focus on the self-managing mechanisms is very limited and is handled in an adhoc manner based on the customer needs or the defects that come post release of the software [4]. The other major advantage is the automation that is provided by the controllers. They come with a mathematical model that justifies their stability, convergence and accuracy [4]. With the advent of the internet and changes that are occurring to businesses and their service delivery modes, there is a big necessity for the computing platforms and systems to operate in self-healing and self-managing manner. Control theory has become an important choice to provide these kinds of capabilities in the computing environments.

Most of the software application layers including enterprise and cloud applications use Web and Application Servers as platform to host the service delivery components [5]. Enterprise Service Bus is heavily used in Service Integration both in enterprise and cloud environments [6]. It is highly desirable that high performance and scalability [7] to be provided by these software applications. This triggered the researchers in investigating the applicability of control systems in Web, Application Servers with some interesting results.

We attempt to study control systems application in web and application servers and provide a useful review for either researchers or practitioners of this subject. This paper is a first step towards investigating how the control systems theory is applied to the problem areas in Enterprise and Cloud environments that employ Web Servers, Application Servers and Enterprise Service Bus. We have not discussed the experimental results achieved in the various research papers but the conclusion is highlighted as required. The figures are referred from the original papers and the references are appropriately cited. Our study is based mainly on Java based Web and Application servers as they are one of the successful platforms that host applications for service delivery either enterprise or cloud environments.

The paper is organized in the following manner. Firstly, we discuss the background of why control system is suitable for solving some of the challenges in computing environments followed by a brief overview of control system concepts. In the next set of sections we discuss how different type of controllers are being applied in Java based Web, Application Servers, and Enterprise Service Bus and Cloud environments. We conclude with our observations, possible future research areas and our next steps in this research.

II. Background

Feedback Control Systems has been in practical application in many engineering disciplines [8]. The inherent characteristics due to the feedback phenomenon bring significant self-managing capabilities hence it is a successful theory. It brings strong mathematical concepts to model both linear and non-linear systems [9], and determine the stability and controllability of the system using techniques such as Nyquist [10], Routh-Horwitz Criterion [11]. The Control theory provides pre-defined controllers such as P, PI, and PID

Controllers. The rich theory has motivated the researchers to start experiments in using control systems in computing. The advantage of having some basic controllers is put to use in improving the performance of computing systems. In the case of distributed systems control systems is mainly used to improve performance of web or application servers by improving their caching or tuning other important parameters such as cpu, memory utilization, no of users. The study shows the feedback control system efficient in building self-managing middleware systems such as CORBA where a control service (FC-ORB) was implemented to improve the CPU utilization [12]. It is interesting to observe discrete control systems achieving IT workflow automation, easing the IT management [13]. In the case of computer networking the controllers are put to use for flow and congestion control [14][15]. There are additional applications in Database and DatawareHouse servers to improve the memory management [16]. It is interesting to note that the majority of investigation of control systems in computing environments is in distributed computing system environments [17][18][19][20] and computer networking [21][22][23][24]. We observe that feedback control mechanisms though have been well explored; still there is scope to extend the research in making it as a first class building block of the distributed environments. Additionally the recent trends in the Enterprise and cloud environments [25] demand more self-managing capabilities, which motivated us to conduct a study on this subject.

III. Control Theory Primitives

In this section we discuss types of controllers, control strategies and modeling of control systems theory briefly. The Figure 1 shows a typical feedback control system where the controller is used to tune the system to be controlled in order to achieve the desired output. The system will have external disusturbance also called noise that affect the system behaviour.

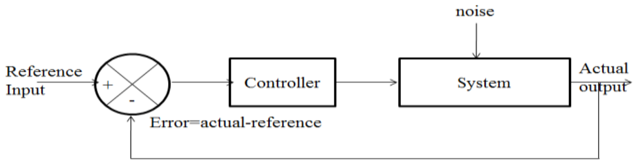


Figure 1. A typical Feedback Control System

A. *Controllers* [25]

- Proportional (P): The controller output is proportional to the error signal.
- Proportional-Integral (PI): The controller output is dependent upon the proportional and integral values to the error signal.
- Proportional-Integral-Derivative (PID): The controller output is dependent upon three separate constant parameters, proportional derivative and integral to the error signal.

B. *Control Strategies* [26]

- Hierarchical: In this type of control strategy, the governing control software is arranged in a tree. If the tree links are implemented by computer network then it is a Networked Control System.
- Robust: Robust control works within pre-defined boundaries of a data set. It is suitable in the presence of bounded modeling errors.
- Adaptive: Adaptive control involves in modifying the control law when the system parameters to be controlled are time-varying or uncertain in nature. It is widely used in self-tuning of the system to be controlled.
- Stochastic: Stochastic control is used when there is uncertainty in the data. It designs optimal control that performs the desired control task with minimum average cost despite presence of the noises.
- Optimal: Optimal control deals with finding control law of a system to achieve a certain optimality criterion are achieved. It is a set of differential equations to minimize the cost functional.
- Intelligent: The controllers that use AI computing techniques like neural network, Bayesian probability, fuzzy logic, machine learning, and genetic algorithms come under Intelligent Controllers

C. *Modeling Computing Systems*

The following are the various modeling techniques used for representing the systems on which the control techniques are applied.

- Queuing Theory: It is a mathematical model of queues. It helps in calculation of performance measures of the system to be controlled.[27]
- Model Predictive Control(MPC): MPCs [28] rely on dynamic models of the process, most often linear empirical models obtained by system identification
- Petri Nets: Petri Nets [29] are popular mathematical model languages to represent Distributed Systems. It is a bipartite graph. The nodes represent transitions and places. The directed arcs describe which places are pre-and/or post conditions for which transitions.

D. *Types of Control Systems* [30]

- Linear Systems are mathematical models that have the properties of superposition and homogeneity.
- Non-Linear systems do not satisfy the properties of superposition and homogeneity.

IV. Web Servers

In this section we discuss the application of control systems in Web environments such as web servers and web services.

A. *Web Caching*

With the advent of internet the importance of Web Servers is very significant and it is obvious that they run with high performance always. There are many ways to improve the web server's performance, Web caching being an important measure. In Web Servers it is desirable to self-manage the web cache in high traffic internet environments. Web proxy Cache provides a key differentiation in the information access on web servers and performance differentiated architecture in proxy caches is proposed in [31]. The content is classified into multiple classes to cache in the web proxy. It proves the average response time of client decreases using a Hit Rate Control Loop implemented over Squid proxy cache. A feedback mechanism is used along with space allocation heuristic to adjust the cache space based on the actual (Ri) and desired(R) relative hit ratios. The Figure 2 below illustrates the controller designed to minimize the error. The figure is redrawn from the original paper to better clarity.

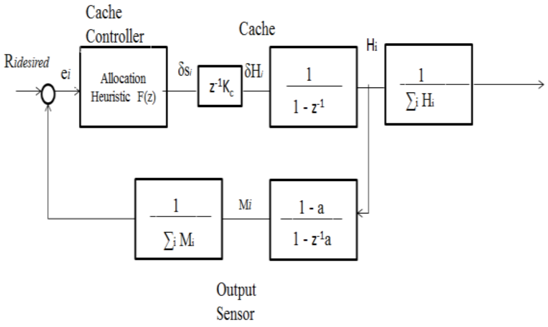


Figure 2. Web Caching Feedback loop (in z-transform)

But this method has the drawback of fixed parameters in the model that are used to adjust the cache size. Additionally, empirical and synthetic workloads are used for experimentation. The uncertainty of the parameter estimate is not incorporated in the controller design proposed in [31].

An improvement is proposed in [32] as shown in Figure 3 which is an adaptive control technique used. The controller design with on-line recursive parameter estimation for QoS guarantees in distributed environments is implemented. Additionally the parameter estimate is incorporated in the controller design. The empirical and synthetic workloads are used are improved for effective convergence of online estimates of the parameters in [32] for the stochastic workloads are considered.

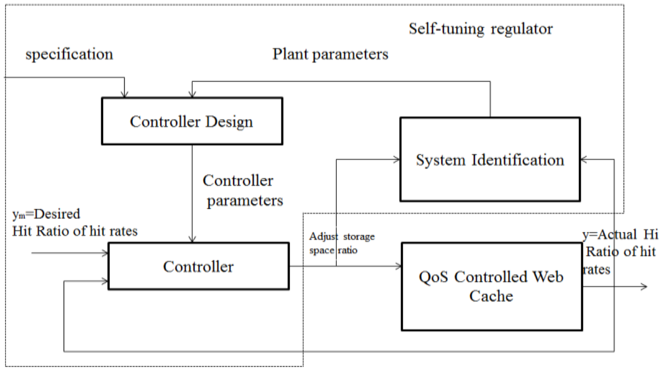


Figure 3. Adaptive QoS Web Cache control system

To summarize, in the adaptive cache control in [32], the QoS of the web cache is managed by including the sensitivity analysis which has the following factors: excitation signal, workload uncertainty and a priori knowledge of the system.

$$u(k) = \sum_{j=1}^{i} a_{i\text{-}j}(k)u(k-j) + \sum_{j=1}^{i} b_{i\text{-}j}(k)[y_m(k-j) - y(k-j)] \quad (1)$$

$a_{i\text{-}j}(k)$ and $b_{i\text{-}j}(k)$ are controller parameters. At the end of each sampling time the controller is fed with output y reference y_m and input u. The controller order is represented by l. The output should track asymptotically the reference if the estimates are accurate. But the conventional adaptive control has a dilemma between asymptotically good control and asymptotically good parameter estimate. Additionally for uncertain and dynamic work load large scale distributed environments, it is difficult to build good a priori knowledge. This constraint reduces the prediction accuracy and increases time for prediction to converge. This uncertainty is handled by implementing a dual control framework as shown in Figure 4. Which is a redrawn based on the original paper for clarity purposes.

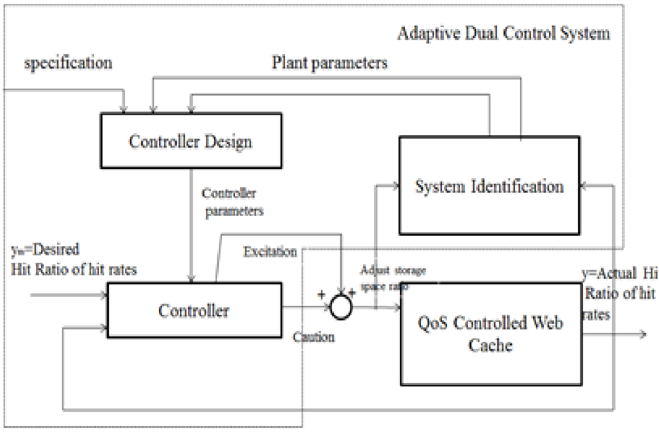


Figure 4. Adaptive QoS Web Cache dual Control

It incorporates uncertainty in the control strategy with the control signal. The system modeled is a discrete time-varying as in Equation 1. The Equation 2 below represents the hit rate at k + 1 interval.

$$y(k+1) = -a_1(k)\, y(k) + ..-a_n(k)\, y(k-n+1) + = -b_1(k)\, u(k) + ..-b_m(k)\, u(k-m+1) \quad (2)$$

where

$y(k)$ - actual ratio of hit rate
$u(k)$ - control signal for adjusting storage space ratio
k - discrete time index

$a_i(k)$ and $b_j(k)$ for $i = 1..n$ and $j = 1..m$ are unknown time-varying system parameters. The uncertainty is modeled using an additional stochastic parameter drift in the Equation shown below:

$$\mathbf{p}(k+1) = \mathbf{p}(k) + \varepsilon(k) \quad (3)$$

where

ε(k) - is the white noise drift vector.

p(k) is estimated using standard technique for online system identification.

In order to derive control law the following cost functions (4) and (5) have to be minimized. These equations will help in meeting the goals of dual adaptive controller to control the system output and to accelerate the estimation for future control improvement.

$$J^{a}{}_{k} = -\mathrm{E}\{\beta^2[y_n(k+1) - y(k+1)]^2 \mid \mathfrak{I}_k\} \qquad (4)$$

Where

$\mathfrak{I}_k$ = represents the set of inputs and outputs at time k

β^2 = coefficient to simplify algebraic manipulations

$y(k + 1)$ = output

The equation (4) is used to minimize the deviation from system output $y(k + 1)$from nominal output $y_{\mathrm{n}}(k + 1)$

$$J^{a}{}_{k} = -\mathrm{E}\ \{\ [\ y(k+1)] + \sum_{i=1}^{m} c_i\, y(k - \mathrm{i} + 1) - \mathrm{p}^{\mathrm{T}}\mathrm{m}(k)]^2 \mid \mathfrak{I}_k\} \qquad (5)$$

c_i = desired pole values

p = predictive error value

The equation (5) is used to accelerate the parameter estimation. We infer that the dual controller framework optimizes the tradeoff between the control goal and the uncertainty prediction.

The Figure 5 shows how adaptive control law is applied to improve the web cache hit ratio discussed in [33]. Two adaptive controls are designed: a deterministic control to deal with the parameter uncertainities and a stochastic design to compenstate system noises. The cache is classified a multiple classes of content in proxy cache. In their empirical study it is observed that the ratio of average hit ratio is closely related to the ratio of cache storage assigned to the classes. The experimental results prove stochastic adaptive controller performs better than a deterministic controller.

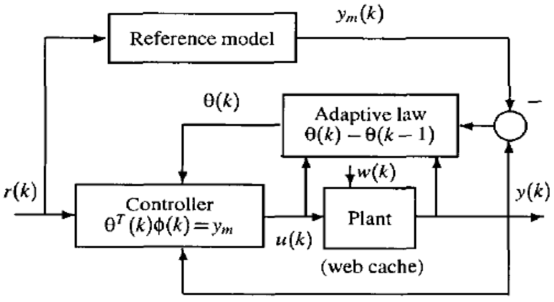


Figure 5. Direct adaptive control system for a Web Cache

B. Web Server Performance

In this section we discuss on improvement of the performance of Web Servers. In [34] a general control theoretic model is developed and validated on a general single server queue. In [35] the authors have modeled the system using linear MIMO of Apache Web Server to design feedback controllers, analyze pure pole placement and LQR [36] based techniques which doesn't have the imaginary part in the pole placement. This was one of the initial works on control theory on computing. The focus is on regulating CPU and Memory utilizations identifying tuning parameters for controller design as shown in Figure 6. The tuning parameters are "KeepAliveTimeOut" (KA) and "MaxClients" (MC). Also the interest is to modify these parameters dynamically. An empirical approach is used to model the server based on the statistical models of the data which is a linear time invariant model (ARX) in nature. The proposed PI controller in [35] aims at managing between speed of response and overreaction to noise through PI controller. In their pole placement technique it is observed that large control gains result in excessive control reaction to the stochastic of the system and large changes in KA and MC leading more oscillatory closed loop response. Their experiments proved that LQR controller has smaller gains leading to a less oscillatory behavior. The equation (6) represents the LQR low gain controller.

$$\mathrm{J} = \sum_{k=1}^{\infty}\ [\,e_k\ v_k]^{\mathrm{T}}.Q.\begin{bmatrix} ek \\ vk \end{bmatrix} + u^{T}{}_{k}R.u_k \qquad (6)$$

As this is the initial works in a MIMO model, a limited Tuning or control parameters are considered. Usually the linear models are proper for good operating regions.

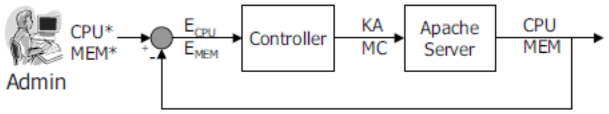


Figure 6. Feedback control of Apache Server for CPU and MKemory Utilizations

But the practical environments exhibit non-linear behavior. A non-linear control theory is discussed in [37] to design admission control mechanisms of a server system which is modeled as a GI/G/1-system. The related work in [38] uses PI controller in admission control of servers using linear control theory. Though PID controllers are used in ATM flow control discussed in [39] they are explained using linear deterministic models and in [40] it is discussed such models cannot be used in queuing systems. So a non-linear control is considered and a PI and RST controllers are used in [41] where the server requests are modeled as queuing system. The admission control has a gate, a controller and a monitor measuring average server utilization represented by $\rho(kh)$. The control time is divided into 'k' intervals and length of each interval is 'h' seconds. The Figure 7shows the discussed control theoretic model and the system under study is shown in Figure 7. This figure is redrawn from the original picture.

The arriving requests are only admitted if there is an available token. New tokens are generated at the rate of $u(kh)$.

The server utilization during interval kh, $\rho(kh)$ can be estimated as

$$\rho(kh) = \min(\ (u(kh) + x(kh))/\ (\sigma_{max(kh)}),\ 1\ ) \qquad (7)$$

where

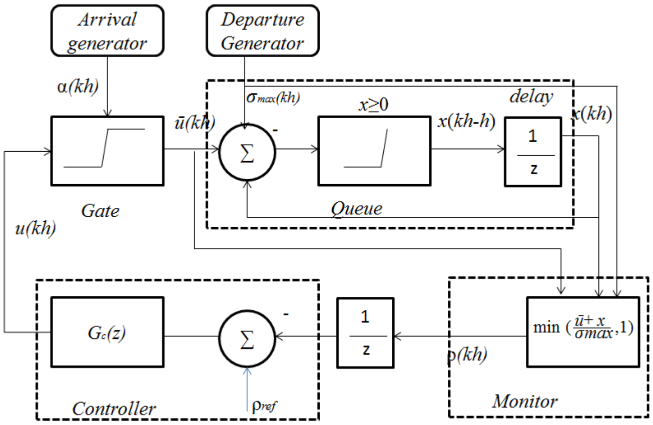


Figure 7. A control theoritic model of a GI/G/1-system with admission control

$u(kh)$ = desired admittance rate for interval kh

$x(kh)$ = represents the state of the queue

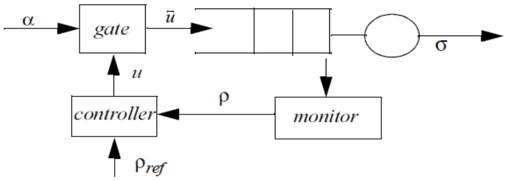


Figure 8. Investigated System

The PI-Controller is represented as shown in the equation (8)

$$u(kh) = K_e(kh) + \sum_{i=0}^{k-1} (K/T_i)\, e(ih)^T \qquad (8)$$

The gain K and T_i integral time are set to make the controlled system behave as desired and need to be determined with respect to stability and robust ness. This control is applied on D/D/1-system, M/M/1-system and M/H_2/1-system and the variations are calculated. Also, the controller was implemented in the discrete-event simulation program and the results are found to be similar to the Simulink model.

Another controller designed is the RST Controller represented as

$$R(q)\, u(kh) = T(q)\, \rho_{ref}(kh) - S(q)\, \rho(kh) \qquad (9)$$

$R(q)$, $T(q)$ and $S(q)$ are expressed in forward shift operator q. They are the controller polynomials and the results between using PI-Controller and RST-Controller; it is observed that the settling time is shorter for the latter. With PI-Controller the bursty stochastic processes are difficult to control and it becomes important to consider the non-linearity and stochastic of the systems to be controlled and model the systems in that manner. The objective is to keep the server utilization as close to reference value and the settling time should be short. The stability of the server node is analyzed when a PI-controller is applied based on linear queue model, and compared with the admissible control parameters derived from nonlinear analysis. In order to study stability Tsypkin criterion [42] or the Jury-Lee criterion [43], which are the discrete counter parts of the popov criterion, continuous systems are used.

SLA management of Web Servers is managed using "Queue Length Model Based Feedback Control" under dynamic traffic that provides delay regulation reducing residual errors, which is discussed in [44] as shown in Figure 9.

The queue length is used for controller design related to the delay of a request. The queue length is predicted at each control invocation and updates the request rate estimate. Based on the new queue length based feed forward predictor adjusts the estimated service rate based on request rate estimation and server queue built up. The controller calculates the service rate adjustment based on difference between delay reference and measured delay in each control interval.

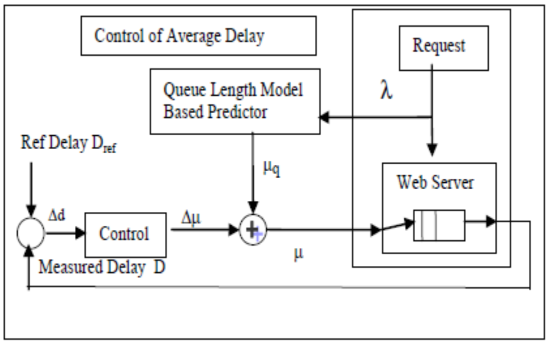


Figure 9. Queue Length Model Based Feedback Control Architecture

The classic Queuing based feedback control is observed to be ineffective with bursty traffic [45] and a Pareto On/Off distribution model [46] is used to model such web traffic. It is proved that a PI controller applying Queue Length Model based Feedback has better improvement. Queue length is the predicted unlike the load metrics compared to other theories.

We observe PI controller being commonly used to improve the web caching and Web server performance, also adaptive dual controllers designed to improve the performance in web server environments.

V. WEB SERVICES

In this section we discuss how the control system concepts are applied in Web Services environment.

A. *Service Execution Engine QoS*

We have seen the application dual feedback control loops in Web caching [31]. A similar approach can be seen on a Service Execution Engine to provide response time Guarantee even when the workload varies significantly, and increasing the throughput of the engine in WebJetFlow [47] as shown in Figure 10, which is redrawn from the original paper. Most of the workflow engines executing the composite service workflows never focused on the qualities of the composite services in running environment [48]. The solution proposed in WebJetFlow tries to address this concern by classifying the services and assign them to the process executor according to their class using a dual feedback control. The service invocations are assigned the threads according to the class. For each such class there is a queue maintained. Once again a PI – controller is used to control the input which is shown below

$$u_k(t) = u_k\,(\,t - 1) + g.\,(e(t) - r.\,e(t-1)) \qquad (10)$$

where

$e(t)$ is the controller input,

u_k is the controller output,

g and r are design parameters.

The controlled process executor is modeled as a difference equation as

$$v(t) = a_j(t-1) + b_j(t-j) \qquad (11)$$

The root-locus technique is used to place the closed-loop poles for a given response time control by setting g and r.

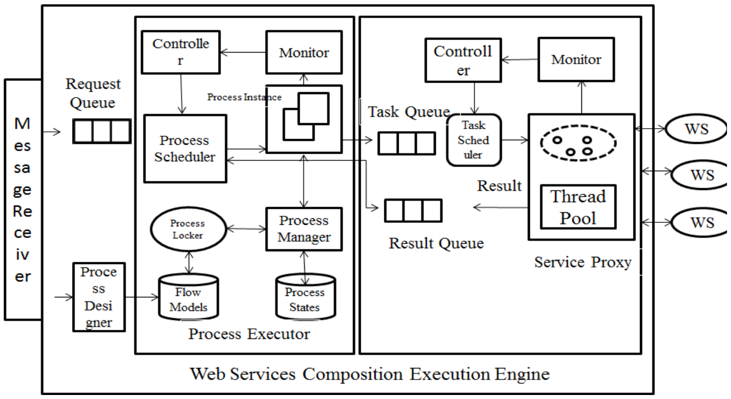


Figure 10. Web Services Composition Execution Engine Framework

B. Adaptive Control in Web Services

ControlWare is developed in [49] which is a middleware package to control the performance and QoS of the Web Servers along with the scheduling and queuing theory. It considers the classical problems of the server performance: rate, delay and ratio control problems.

The delay and ratio pose non-linearities in the system and feed forward control models to predict the system behavior when the inputs are changed. Another important improvement suggested is to retrofit the controller in a non-intrusive approach. The ControlWare provides a means of plug-and-play sensors, actuators, and controllers into performance loops.

The following Figure 11 shows the computing model and Figure 12 shows the control oriented representation of the ControlWare.

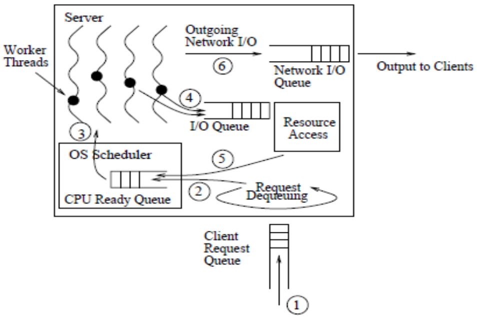


Figure 11. Computing model

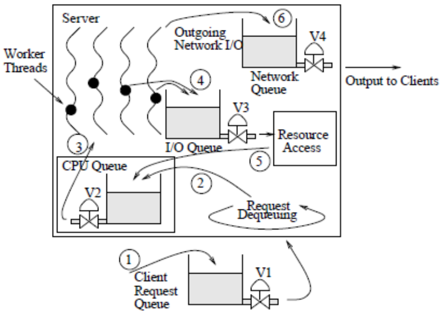


Figure 12. Control Oriented representation

The server is modeled using the ARMA difference equation [50]. The equation can be derived from a state space representation of the server model

$$\mathbf{x}(k) = A\,\mathbf{x}(k\,\text{-}1) + b\,u(k) \qquad (12)$$

$$y(k) = C\,\mathbf{x}(k) \qquad (13)$$

where

$\mathbf{x}(k)$ – state vector,

A,b and C represent system model

The prediction part of the feedback system on controlling the web server performance takes the non-linear behavior of the control loops into account.

VI. Application Servers

The Application Servers provide infrastructure for development and deployment of 3-tier applications. These servers are popular in hosting many mission critical applications and it is important that Application Servers exhibit high performance always. The Application Servers that we refer in this review are Java Enterprise Edition (JEE) [51] based implementations. The previous versions of JEE servers are also called as J2EE Servers. The following discuss the control theory application in different aspects of the Application Servers.

A. *Data Flow Control in J2EE Servers*

The flow control in J2EE servers is discussed in [52] using control theory and dynamic probabilistic scheduling. The web tier load is regulated to prevent the overload on the database tier. The classical control theory principle is supplemented with the workload classification and dynamic queue scheduling to maintain the fairness in the requests, handle larger variance in resource demands. The feedback control is used to regulate the load that enters at the web tier, based on the load at database tier. This is similar to congestion control in networking with variations like support continuous processing, throughput is lower. A simple Integral Controller is used and the control law for *k*th interval is

$$u(k) = u(k - 1) + K_i\, e(k) \qquad (14)$$

$$e(k) = y(k) - r(k)$$

$u(k)$= Number of allowed HTTP request per time period.

$y(k)$ = measured output

$r(k)$ = Maximum database connection per time (reference value)

$e(k)$ = Control error

The database usage is profiled offline and the requests are classified using k-means algorithm as large and small requests. The average usage of the database of each class is passed to the feedback control to estimate the usage for the next period. In order to support small requests Dynamic Probabilistic scheduling (DPS) is introduced by increasing their priority than large requests but providing a fair priority to the large requests. The Figure 13 shows the overall software architecture.

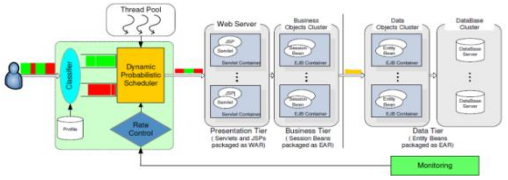


Figure 13. Admission Control Subsystem

The authors used simple Integral Controller to regulate the load on the database tier. But the database usage is classified offline. The controller works like a knob to control incoming requests at the web tier itself. They claimed to improve the accuracy, but this is a simple technique.

B. *Repair Management in J2EE Servers*

The automatic repair control of J2EE servers using feedback control is introduced in [53]. It is called as JADE in which an autonomous repair management system for J2EE clusters is constructed based on FRACTAL model. The feedback control is used by connecting the managed system and the repair management functions. There are multiple features in [53] but we explain only the Repair Management component control loop that uses the feedback control. The Manager component contains a policy component that decides on the node allocation to deploy the application components during repair management. Nodes are referred to as physical computer abstraction. The block diagram is shown in Figure 14.

The sensors monitor the managed system like state, resource usage, and failures. The actuators perform life cycle and configuration actions. The Transport component binds sensors and actuators. The Manager component implements the control loop for analysis and decision of the repair management when there is a failure either in the node or a component in the node.

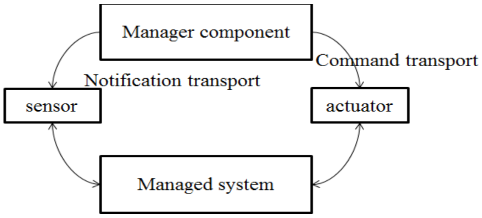


Figure 14. Management Control Loop

Though it is interesting to see the fault and repair management using feedback control, the policies considered are simple and the feedback controller types are not put to use and the modeling and analysis provided by the control theory is not applied completely.

C. *Improving J2EE Server Performance*

In this section we will discuss on how the adaptive control is put to use in the J2EE Servers. In [54] the authors implemented feedback control for adaptive self-configuration capability to improve the performance of the J2EE Servers. The data is collected by the Data Collector which is further processed which in turn is stored in the Database. The predictor will predict the future performance data to pro-actively determine the time-varying future loads using the database. The Comparator compares the difference between the predicted performance data and the SLA data. The Decision Maker will take an appropriate decision on the tuning strategy from the knowledge based which will be applied back to the J2EE Server, based on SLA, predicted values. An experiment is conducted on JBoss Server where it is observed that the EJB Container performance is depending upon the thread pool size (MaxPoolSize). The effect of MaxPoolSize response time is verified

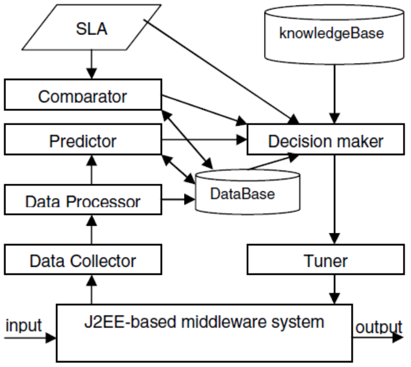


Figure 15. Adaptive performance configuration architecture

An architectural approach is proposed to handle variable workload in [54] called as Adaptive Performance Configuration architecture. But the improvement observed in [55] is to feed measures into the model and adaptively self-configure the system as shown in Figure 15. A knowledge based framework is proposed in [55] that has monitoring and knowledge based configuration mechanisms which is an improvement over [56][57] where linear or feed forward controllers are used. Though there are adaptive controllers proposed in [58] but are limited rule of thumb. The following Figure 16 shows how the adaptive performance is achieved using qualitative knowledge. The knowledge is represented in the form of fuzzy logic. This controller is based on fuzzy control as shown in the Figure below:

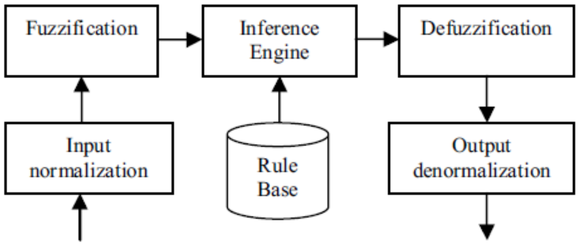


Figure 16. Structure of Fuzzy control

The fuzzy rules are defined for the “Max Pool Size” in JBoss Application Server. The authors of [55] claim that these rules can be proved by Lyapunov method that converge to the optimal MaxPoolSize at steady state. For ease of computation Triangular Membership functions are chosen and Centre of Gravity (COG) is used for defuzzification. The fuzzy control algorithm is used to compute the Next-Change-In-MaxPoolSize for given the values of Change-In-MaxPoolSize and Change-In-Response Time.

We observe that for adaptive mechanisms Fuzzy control is proved to be providing useful results over the conventional controllers.

An automatic tuning system proposed in [55] where controlling of QoS of modern E-Business site is studied. The performance data is monitored regularly and is fine-tuned by a controller if the QoS is not in the acceptable range. The server is tuned to the appropriate configuration to meet the expected QoS. An offline testing is done on the Application Server to identify the best configurations. JBoss Server is chosen to experiment the solution as a target system performing configuration tuning through a Controller, which is shown in Figure 17. This figure is redrawn from the original paper for clarity. There is a monitor component that collects the performance data.

The following are the configuration parameters chosen for performing the tuning Server Throughput, Server Side Response Time (RT). The Controller maintains the Agreed QoS and compares with the Actual QoS periodically. If the evaluation is above fixed threshold then the appropriate server configuration is applied.

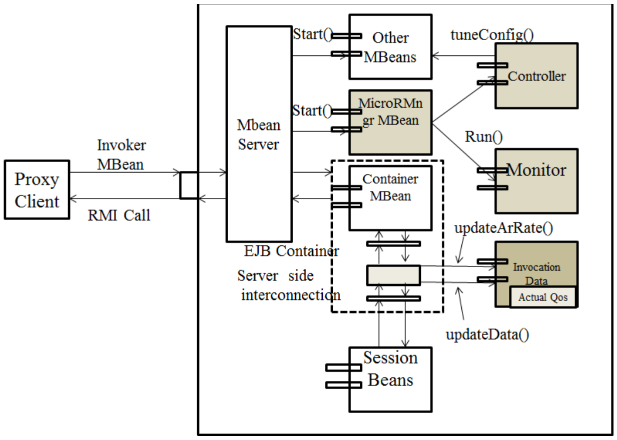


Figure 17. Automated Tuning System

The feedback control system though applied no conventional feedback controller is used. The optimum configurations is decided based on an offline test that is definitely an incomplete exercise. It may not be able to depict the real time load conditions.

Though there are different controllers available, PI controller is being heavily experimented in computing. But the results are showing a trend towards implementing the adaptive and intelligent controls that bring more advantages in improving the performance of Application and Web Servers.

VII. Enterprise Service Bus

Enterprise Service Bus (ESB) has become an important choice for Service Integration in huge enterprise applications. ESB implementations are run as standalone applications, but preferred to run as an integral part of the standard Application (JEE) Servers. A feedback control algorithm is discussed in [59] to realize load balancing routing and fail-over in ESBs that integrate IT services and Telecom services. A Hybrid Services Execution Environment (HSEE) is proposed which is a distributed architecture using ESB framework and routing that has to be available and manage load changes. There are two controllers, local and distributed which work together to maintain load balancing across the ESB nodes. The Figure 18 shows the load balancing using distributed feedback control. The detailed analysis and results are not discussed in our review.

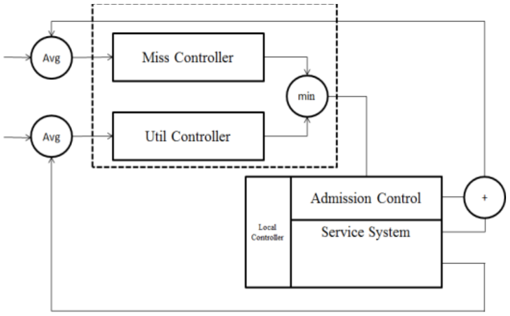


Figure 18. Load Balancing based on Feedback Control

VIII. Cloud Environments and Services

Cloud computing is the most discussed infrastructure platform for many of the service providers to host their application services due to cost based advantages. The service providers avoid investing and maintaining the infrastructure to host their services. In such scenario it is important for the service providers to ensure the QoS and SLA are maintained by them for their end customers. There are variety of cloud service delivery environment developed using JEE platform, besides other choices like OSGi. The control theory is investigated for maintaining the SLA of such cloud services in [60]. Besides the existing challenges and solutions discussed so far, there are additional important problems that need to be addressed in Cloud computing systems. In [61] the authors identified different challenges in providing automated control in cloud environments. The first challenge is to decouple the control into a cloud controller in the cloud infrastructure provider side and the application control to be moved out to the guest. These controllers have independent policies. This requires API from cloud hosting to support the controller's message exchange. The next one is the level of granularity of control. This depends upon how the resource providers choose to export the access to hypervisor level actuators. The previous implementations of the feedback controllers have fine grained access to the sensors and actuators on single virtualized nodes. For horizontally scalable clusters when the granularity is coarse relative to allocated resource, it is required to dampen the control loop at smaller sizes. The API that the Cloud resource provider exposes to the guests need to consider various constraints like how much the internal control to be exposed, how to integrate the guest control policy, how to design effective controller for 3-tier interactions, expose sensors that are suitable for stable control, expose suitable actuators for the controller policy to configure adapt request routing or other programmable network elements. In [61] Proportional Thresholding is proposed to meet these challenges. The solution is to modify P-controller by using a target range, which decreases as the accumulated actuator value increases. To eliminate steady state errors, Integral control is used as a policy defined by the equation (15).

$$u_{k+1} = u_k + K_i(v_{ref} - v_k) \qquad (15)$$

where

u_{k+1} - is the new actuator value

u_k - is the current actuator value,

K_i - is the Integral gain parameter.

v_{ref} *and* v_k are target and current sensor measurements respectively

This solution is experimented with Apache web server was front end, ORCA is used as underlying architecture, Tomcat cluster in the back end. If we observe carefully Tomcat is JEE container and it becomes important to consider the control systems for JEE servers hosting services in cloud environments. In [61] the infrastructure balancing using feedback control is discussed. The challenges when JEE based applications run on such an infrastructure providing cloud based services, the design of controllers for such problems are not discussed in detail. It will be important to investigate the applications of control theory on cloud services that are built using JEE environment.

In [60] the focus is again on the infrastructure layer and discusses briefly the various control theory applications but the objective is on resource allocation mainly. It is evident that large scale resource management problems could be effectively handled by adopting advanced control theory concepts like supervisory, cascaded, hybrid and optimal control.

An autonomic Service Delivery Platform (SDP) is proposed in [62] that host cloud services. We see an opportunity in applying the feedback control theory in building adaptive SDPs.

IX. Conclusions and Future Research

The feedback control theory has inherent advantages as discussed in [2] such as Quantitative I/O models, Dynamics and Transients, Correlation between multiple metrics, control algorithms, stability analysis, non-linear time-varying behavior. These kinds of support elements provide a methodical design and implementation choices of controllers suitable for the computing environments. Though control theory is being applied to various areas in Computing like Networks, Database systems, IT infrastructures, we observed that it is extensively investigated in Distributed Computing Systems starting at elementary components like web caching, Web Services to the Application Servers (e.g., JEE Servers), . We have also noticed that control theory is being investigated in the latest distributed systems like cloud computing and environment that poses a different set of challenges to the controllers like modeling the systems and providing high accuracy in prediction during feedback cycles. We have seen approaches of online modeling of the system to be adapted, applying different controllers starting PI controllers to hybrid controllers to certain extent. But it is essential to mimic the behavior of human in adapting the dynamic varying conditions in the network which can be achieved by combining the dynamic modeling of the systems with hybrid controllers and more innovative controllers, based on the computing system condition. More importantly it is observed that the JEE Servers are not just containers but also moving towards building Service Delivery platforms in SaaS environments, deployed as hosting components for service integration. Though there are very preliminary attempts to apply feedback control, it is essential to investigate it further and propose improved modeling and control designs. The controllers are built as plug and play components in computing environments, but not as obvious as any other plug-n-play architectures. Also the study of applying feedback controls in dynamic module component software like OSGi is not explored. We observe the JEE and OSGi are converging and it may be a good area to explore the feedback control applications in such containers hybrid containers.

Another important relevant work we observe is in modeling. The current modeling of the system for feedback is differential equations based, which brings difficulty in design and implementation of the controller software.

Spring containers have been more popularly in use by many of the distributed software applications at enterprise level. There is an attempt to apply the control theory in spring frameworks [63], but as the spring frameworks applicability is increasing it is obvious that they would face challenges similar to the ones faced by JEE and other Application Servers. It is important to explore the applicability of control theory in spring servers.

Based on this survey we observe that building hybrid intelligent controllers would make feedback controllers for computing, as regular building blocks. Though there are investigations in progress in designing hybrid controllers, most of them are limited to the web server environments [64][65], utilization of the resource management [66][67][68] or server admission control [69].

Data Mining [70] being a confluence of various Machine Intelligence based concepts. It has rich algorithms to handle and process the data efficiently for high accurate prediction capabilities, pattern recognition mechanisms. We are investigating such abilities and how effectively they can be used in designing intelligent hybrid controllers.Improved and hybrid feedback control systems on next generation distributed environments like cloud computing, services and associated technologies would aid in building more robust enterprise and cloud applications and services.

X. FUTURE WORK

We are investigating the applicability of control systems in different areas of the Distributed Systems with a specific emphasis on Java based Enterprise Servers. As a first step towards this we have designed a feedback controller to improve the statement caching mechanism in the JDBC drivers [71]. Our investigations are progressing in identifying optimal controllers for different components in distributed systems.

We are investigating the possibility of bringing the feedback control system modeling into the UML modeling and provide a generic API for integrating feedback controllers with computing systems. This enables the feedback control systems as first class elements of the UML.

Further, we are exploring control system application in Service Orchestration Engines and SDPs that are based on JEE using ESB and OSGi in their environments.Our goal is to design optimal hybrid controllers that are simple and well suited for the latest Java based environments.

REFERENCES

[1] Tarek Abdelzaher, Yixin Diao, Joseph L Hellerstein, Chenyang Lu, and Xiaoyun Zhu., "Introduction to Control Theory and Its Applications to Computing Systems", International Conference on Measurement and Modeling of Computer Systems SIGMETRICS'08

[2] What Does Control Theory Bring to Systems Research? Xiaoyun Zhu, Mustafa Uysal, Zhikui, Wang , Sharad Singhal, Arif MerchantPradeep Padala, Kang Shin, ACM SIGOPS Operating Systems Review, Volume 43 Issue 1, January 2009

[3] Chris Toft et al, www.hpl.hp.com/techreports/2004/HPL-2004-49.pdf Self Managed Systems - A Control Theory Perspective HPL, 2004

[4] Joseph L. Hellerstein, Yixin Diao, Sujay Parekh, and Dawn Tilbury Feedback Control of Computing Systems, John Wiley 2004

[5] Web and Application Servers for services delivery, http://en.wikipedia.org/wiki/Service_delivery_platform

[6] ESB Service Integration http://en.wikipedia.org/wiki/Enterprise_service_bus

[7] Yixin Diao, Joseph L. Hellerstein, and Sujay Parekh "Control of Large Scale Computing Systems", ACM SIGBED Review, Vol3, Issue 6, 2006

[8] I.J.Nagrath, M.Gopal,"Control Systems Engineering,", New Age International Publishers, 2007

[9] Zoran Gajic Lecture notes on Nyquist criterion, "www.ece.rutgers.edu/~gajic/psfiles/nyquist.pdf"

[10] Kunio Takaya,Lecture Notes on Routh Hurwitz criterion, "http://www.engr.usask.ca/classes/EE/481/takaya_notes/ee481-p8-routh_hurwitz.pdf", 2009

[11] Xiaorui Wang et al, "FC-ORB: A Robust Distributed Real-time Embedded Middleware with End-to-End Utilization Control", ACM Journal of Systems and Software, Vol 80, Issue 7, 2007

[12] Yin Wang, Terence Kelly, Stéphane Lafortune, "Discrete Control for Safe Execution of IT Automation Workflows" , ACM European Conference on Computer Systems, 2007

[13] Seungwan Ryu, Chulhyoe Cho,"PI-PD-controller for robust and adaptive queue management for supporting TCP congestion control", 132 - 139 18-22 April 2004

[14] Ohsaki, H. Murata, M. Ushio, T. Miyahara, H, "A control theoretical approach to a window-based flow control mechanism with explicit congestion notification", Pages 2715-2720, vol.3 IEEE Conference on Decision and Control, 1999

[15] Kyoung-Don Kang and Jisu Oh, Sang H. Son," Chronos: Feedback Control of a Real Database System Performance", Real Time Systems Symposium, Pg No.267-276, IEEE 2007

[16] C. Lu, X.Wang, and X. Koutsoukos "Feedback utilization control in distributed real-time systems with end-to-end tasks" IEEE Trans. Parallel Distrib. Syst., 16(6):550–561, June 2005

[17] Baochun Li, Klara Nahrstedt," Impact of Control Theory on QoS Adaptation in Distributed Middleware Systems ",American Control conference, 2987-2991, vol 4, 2001

[18] R. Zhang, C. Lu, T. F. Abdelzaher, and J. A. Stankovic. "ControlWare: A Middleware Architecture for Feedback Control of Software Performance" International Conference on Distributed Computing Systems (ICDCS), IEEE Vienna, Austria, July 2002.

[19] S. Abdelwahed, N. Kandasamy and S. Neema, "A Control-Based Framework for Self-Managing Distributed Computing Systems",Workshop on Self-Managed Systems (WOSS'04), Newport Beach,CA USA, 2004.

[20] S. Mascolo, "Classical Control Theory for Congestion Avoidance in High-Speed Internet", 38th IEEE Conference on Decision and Control, pp. 2709-2714, 1999

[21] Lui Sha, Xue Liu, "Queueing Model Based Network Server Performance Control ", Real Time Systems Symposium, 2002

[22] Jose Yepez, Pau Marti, Josep M.Fuertes, "Control Loop Performance Analysis over Networked Control Systems". IECON02

[23] Phillip M. Dickens ,Vinod Kannan, "Application-Level Congestion Control Mechanisms for Large Scale Data Transfers Across Computational Grids", The International Conference on High Performance Distributed Computing and Applications 2004

[24] Brandic, I., "Towards Self-Manageable Cloud Services", Computer Software and Applications Conference, pages 128-133, 2009. COMPSAC '09.

[25] Zoran Vukic, Ognjen Kuljaca, "Lecture Notes on PID Controllers, http://arri.uta.edu/acs/jyotirmay/EE4343/Labs_Projects/pidcontrollers.pdf,

[26] Control Strategies: http://en.wikipedia.org/wiki/Control_theory

[27] Sheldon M. Ross, "Introduction to Probability Models", Academic Press, 2009

[28] Model Predictive Control (MPCs): http://en.wikipedia.org/wiki/Model_predictive_control

[29] Gianfranco Balbo ,"Introduction to Stochastic Petri Nets" : http://www.mendeley.com/research/stochastic-petri-nets-an-

introduction-to-the-theory/, Lectures on Formal Methods and Performance Analysis 2001

[30] Lecture Notes, “http://www.me.berkeley.edu/ME237/2_general_properties.pdf”

[31] Ying Lu, Avneesh Saxena and Tarek E Abdelzaher Differentiated Caching Services; A Control-Theoretical Approach, IEEE International Conference on Distributed Sysytems, 2001

[32] Keqiang Wu, David J. Lilja, Haowei Bai “The Applicability of Adaptive Control Theory to QoS Design: Limitations and Solutions”, IEEE Parallel and Distributed Processing Symposium, 2005

[33] Ying Lu, Tarek Abdelzaher and Gang Tao, “Direct Adaptive Control of A Web Cache System”, Proceedings of the American Control Conference, Denver, Colorado, 2003

[34] A. Robertsson, B. Wittenmark, M. Kihl, and M. Andersson “Design and evaluation of load control in web server systems”, IEEE American Control Conference, 2004.

[35] N. Gandhi and D. M. Tilbury, Y. Diao, J. Hellerstein, and S. Parekh “MIMO Control of an Apache Web Server, Modeling and Controller Design”, IEEE American Control Conference, 2002

[36] LQR, http://en.wikipedia.org/wiki/Linear-quadratic_regulator

[37] Xue Liu, Jin Heo, Lui Sha, “Modeling 3-Tiered Web Applications”, Modeling, Analysis, and Simulation of Computer and Telecommunication Systems,. 13th IEEE International Symposium, 307 – 310, 2005

[38] C. Lu, T.F. Abdelzaher, J.A. Stankovic and S.H. Son, “A feedback control approach for guaranteeing relative delays in web servers” Proc. of the 7th IEEE Real-Time Technology and Applications Symposium, pp 51-62, 2001

[39] A. Kolarov and G. Ramamurthy, “A control-theoretic approach to the design of an explicit rate controller for ABR service”, IEEE/ACM Transactions on Networking, Vol. 7, No. 5, , pp 741-753, Oct. 1999

[40] S. Stidham Jr., “Optimal control of admission to a queueing system”, IEEE Transactions on Automatic Control, Vol.30, No.8, pp 705-713, Aug 1985

[41] M. Kihl, A. Robertsson, and B. Wittenmark, “Performance modelling and control of server systems using non-linear control theory.” Berlin, Germany: 18th International Teletraffic Congress, Sept. 2003.

[42] Michael Larsen, Petar V. Kokotović, “A brief look at the Tsypkin criterion: from analysis to design”, International Journal of Adaptive Control and Signal Processing, Vol 15, Issue 2, Pages 121-128, Mar 2001

[43] Jury-Lee criterion: http://en.wikipedia.org/wiki/Jury_stability_criterion

[44] Xue Liu, Rong Zheng, Jin Heo, Qixin Wang, Lui Sha, “Timing Performance Control in Web Server Systems”, ACM ICAS-ICNS’05

[45] L. Sha, X. Liu, Y. Lu, T. Abdelzaher, “Queueing Model Based Network Server Performance Control”, IEEE Real-Time Systems Symposium, Phoenix, Texas, December, 2002

[46] Pareto distribution: http://en.wikipedia.org/wiki/Pareto_distribution

[47] Chunming Gao, Weian Chen, Huowang Chen, “A Feedback Control Framework of Service Composition Execution for Response Time Guarantee”, Chunming Gao, Weian Chen, Huowang Chen, IEEE ICWS, 2007

[48] Ivona Brandic et. al.. “QoS Support for Time-Critical Grid Workflow Applications”, e-science, pages 108-115, First International Conference on e-Science and GridComputing (e-Science'05), 2005.

[49] Tarek Abdelzaher. Yina Lu, Ronahua Zhana, Dan Henriksson, “Practical Application of Control Theory to Web Services”, American Control Conference, 2004

[50] ARMA: http://en.wikipedia.org/wiki/Autoregressive_moving_average_model

[51] JEE: http://www.oracle.com/technetwork/java/javaee/tech/index.html

[52] Wei Xu, Zhangxi Tan, Armando Fox, David Patterson, “Regulating Workload in J2EE Application Servers”, http://www.controlofsystems.org/febid2006/files/16225_Wei.pdf

[53] Sara Bouchenak, Fabienne Boyer, Daniel Hagimont, Sacha Krakowiak et al., “Architecture-Based Autonomous Repair Management: An Application to J2EE Clusters”, ICAC’05

[54] Yan Zhang, Wei Qu, Anna Liu, “Adaptive Self-Configuration Architecture for J2EE-based Middleware”, Vol 9, HICSS’06

[55] Giovanna Ferrari, Santosh Shrivastava,Paul Ezhilchelvan, “An Approach to Adaptive Performance Tuning of Application Servers”, IEEE International Workshop on QoS in Application Servers, 2004

[56] N. Gandhi, J. L. Hellerstein, S. Parekh, and D. M. Tilbury, “Managing the Performance of Lotus Notes: A control TheoreticApproach”, in Proceedings of 27th International ComputerMeasurement Group Conference, 2001

[57] L. W. Russell, S. P. Morgan, and E. G. Chron, “Clockwork A new movement in autonomic systems”, IBM SYSTEMSJOURNAL, VOL 42, NO 1, pp77-84, 2003

[58] M. Raghavachari, D. Reimer, and R. D. Johnson, “The Deployer’s problem: Configuring Application Servers for Performance and Reliability”, In Proceedings of the 25th International Conference on Software Engineering (ICSE’ 03), 2003

[59] Yang Zhang, “Dependable ESB Routing in Hybrid Service Execution Environment”, AISS : Advances in Information Sciences and Service Sciences, Vol. 2, No. 1, pp. 83 ~ 93, 2010

[60] Harold C. Lim, Shivnath Babu, Jeffrey S. Chase, Sujay S. Parekh, “Automated Control in Cloud Computing: Challenges and Opportunities Challenges and Opportunities”, ACDC’09

[61] Christos A. Yfoulis and Anastasios Gounaris, “Honoring SLAs on cloud computing services: a control perspective”, Proceeding of the 2nd workshop on Bio-inspired algorithms for distributed systems,Pages:29-38, 2010

[62] Robert D. Callaway, Michael Devetsikiotis, Yannis Viniotis, Adolfo Rodriguez, “An Autonomic Service Delivery Platform for Service-Oriented Network Environments”, vol. 3 no. 2, pp. 104-115, April-June 2010

[63] Dr. Wolfgang Winter , “Applying control theory concepts in software applications”, http://www.theserverside.com/feature/Applying-control-theory-concepts-in-software-applications

[64] Yaya wei, Chuang Lin, Thiemo Voigt, Fengyuan Ren, "Fuzzy Control for Guaranteeing Absolute Delays in Web Servers, QoS in Wireless Networks" International Conference

[65] Jianbin Wei ,Cheng-Zhong Xu, “eQoS: Provisioning of Client-Perceived End-to-End QoS Guarantees in Web Servers”, IEEE Trans on Computers, 2006

[66] Mehmet H. Suzer ,Kyoung-Don Kang, “Adaptive Fuzzy Control for Utilization Management”, ISORC , Pages 383-390, 2008

[67] Palden Lama , Xiaobo Zhou , “Autonomic Provisioning with Self-Adaptive Neural Fuzzy Control for End-to-end Delay Guarantee”, IEEE International Symposium on Modeling, Analysis and Simulation of Computer Telecommunication Systems, 2010

[68] P. Lama , X. Zhou, “Efficient server provisioning for end-to-end delay guarantee on multi-tier clusters. In Proc. IEEE Int’l Workshop on Quality of Service (IWQoS), 2009

[69] Jiang Ying, , Meng Dan, “Enforcing Admission Control, Using Admission-Time-Ratio and PI Controller,” Journal of Computer Research and Developmen, 2007

[70] Jiawei Han, Micheline Kamber, , “Data Mining – Concepts and Techniques”, Morgan Kaufmann Publishers, 2006

[71] Ravi Kumar Gullapalli, Dr.Chelliah Muthusamy, Dr.A.Vinaya Babu, Raj.N.Marndi “A Feedback Control Solution in improving Database Driver Caching”, IJEST,Vol 3, No 7, Jul 2011

AUTHORS PROFILE

Ravi Kumar Gullapalli is working as aTechnical Expert in Hewlett-Packard., Bangalore, India. He obtained his M.Tech in Computer Science from Birla Institute of Technology, Mesra,India. He is currently pursuing Ph.D from JNTU Hyderabad,AP, India.

Dr.Chelliah Muthusamy is Academic Relations Head at Yahoo, Bangalore,. He obtained his Ph.D from Georgia Tech and M.Sc(Engg) in Computer Science from Indian Institute of Science(IISc), Bangalore India

Dr.A.Vinaya Babu is a Professor of Computer Science working as Director, Admissions at JNTU, Hyderabad, AP, India. He obtained his Ph.D and M.Tech in Computer Science from JNTU, Hyderabad.

Multimodal Optimization using Self-Adaptive Real Coded Genetic Algorithm with K-means & Fuzzy C-means Clustering

Vrushali K. Bongirwar
IVth semester M.Tech., CSE,
RCERT., Chandrapur(m.s.) India

Rahila Patel
RCERT, Chandrapur (M.S.)
India Pin: 442403

***Abstract*— Many engineering optimization tasks involve finding more than one optimum solution. These problems are considered as Multimodal Function Optimization Problems. Genetic Algorithm can be used to search Multiple optimas, but some special mechanism is required to search all optimum points. Different genetic algorithms are proposed, designed and implemented for the multimodal Function Optimization.**

In this paper, we proposed an innovative approach for Multimodal Function Optimization. Proposed Genetic algorithm is a Self Adaptive Genetic Algorithm and uses Clustering Algorithm for finding Multiple Optimas. Experiments have been performed on various Multimodal Optimization Functions. The Results has shown that the Proposed Algorithm given better performance on some Multimodal Functions.

Keywords*—*Genetic Algorithm (GA); self-adaptation; Multimodal Function Optimization; K-Means Clustering; Fuzzy C-Means Clustering.

I. INTRODUCTION

Optimization means solving problems in which one seeks to minimize or maximize a real function by systematically choosing the values of real or integer variables from within an allowed set.[5][12]

In Genetic algorithm (GA), we not only need to choose the algorithm, representation and operators for the problem, but we also need to choose parameter values and operator probabilities so that it will find the solution efficiently. This process of finding appropriate parameter values and operator probabilities is a time consuming task and considerable effort has gone into automating this process. Literature contains various studies on setting of strategy parameters [1], [2], [3]. There are various types of Optimization functions as unimodal, multimodal, etc. and GA is used to optimize them.

In real-coded GA (RCGA) chromosome is a real-parameter decision-variable vector. The recombination operation is a method of sharing information among chromosomes [8][9].Traditional real-coded genetic algorithm (RCGA) has parameters that must be specified before the RCGA is run on a problem. The most important parameters are the strategy or control parameters: population size (N), mutation rate (pm), crossover rate (pc), and probability distribution index (η).[4][11]

Multi-modal functions have multiple optimum solutions, of which many are local optimal solutions. Multi-modality in a search and optimization problem usually causes difficulty to any optimization algorithm in terms of finding the global optimum solutions. Therefore, when dealing with multi-modal functions, some modification is necessary to the standard GAs to permit stable subpopulations at all peaks in the search space.

In this paper, we have described a framework of GA for multimodal optimization which use clustering algorithm to identify different region of attractors, then search each attractors independently. For generating subpopulations we have used two different clustering techniques K-means and Fuzzy C-means.

Paper is organized as: section 2 introduces multi-modal optimization & Literature for solving them. In section 3 we introduced Clustering, K-means and Fuzzy C-means clustering. In section 4 Proposed Algorithm i.e. Self Adaptive GA with Clustering is described. In section 5 experimental setup is discussed. Section 6 presents empirical results of Proposed Algorithm over few multi-modal test functions. Finally in section 6 we draw some conclusion.

II. INTRODUCTION TO MULTI-MODAL OPTIMIZATION & LITERATURE

As the name suggests, multi-modal functions have multiple optimum solutions, of which many are local optimal solutions. Multi-modality in a search and optimization problem usually causes difficulty to any optimization algorithm in terms of finding the global optimum solutions. This is because in these problems there exist many attractors for which finding a global optimum can become a challenge to any optimization algorithm. In the case of peaks of equal value (height), the convergence to every peak is desirable, whereas with peaks of unequal value, in addition to knowing the best solutions, one may also be interested in knowing other optimal solutions. Therefore, when dealing with multi-modal functions, some modification is necessary to the standard GAs to permit stable subpopulations at all peaks in the search space. [13]

If a classical point-by-point approach is used for this purpose, the method must have to be used many times, every time hoping to find one of the optima.

Various techniques are adopted to optimize multimodal functions. Genetic Algorithm is used in combination with niching techniques. A niching method must be able to form and maintain multiple, diverse, final solutions, whether these solutions are of identical fitness or different fitnesses. A niching method must be able to maintain these solutions for a large enough iterations.[13]

Petrowski suggested the clearing procedure, which is a niching method inspired by the principle of sharing of limited resources within subpopulations of individuals characterized by some similarities. A clustering algorithm [13][17] is used to divide the population into niches. Crowding methods insert new elements in the population by replacing similar elements.[18] Harik [19] introduced a modified tournament algorithm that exhibited niching capabilities.

Species conservation techniques are used to identify species within a population and to conserve the identified species in the current generation..[10]

Multimodal Functions have multiple local as well as global optimas. Any standard GA is found to be less efficient for the various multimodal as well as multimodal multipeak functions, because during execution when the algorithm finds first optima it get terminated with the single optima as result. But for Multimodal as well as multimodal Multipeak functions, the aim is to search all the optimas present in the given function. For this reason, the Initial Population must be divided into clusters. We are assuming, in each cluster we will find an optima. So we can reach to all the optimas.

So we have used K-means as well as Fuzzy C-means Clustering for dividing the Population into clusters.

III. CLUSTERING K-MEANS & FUZZY C-MEANS CLUSTERING

Clustering is the unsupervised classification of patterns (or data items) into groups (or clusters). A resulting partition should possess the following properties: (1) homogeneity within the clusters, i.e. data that belong to the same cluster should be as similar as possible, and (2) heterogeneity between clusters. i.e. data that belong to different clusters should be as different as possible. Several algorithms require certain parameters for clustering, such as the number of clusters and cluster shapes.

Let us describe the clustering problem formally. Assume that S is the given data set:

$S = \left\{ \vec{x}_{1,} \ldots\ldots \vec{x}_n \right\}$ where $\vec{x}_i \in R^n$ .The goal of

clustering is to find K clusters $C_{1,}$ $C_{2,\ldots\ldots}$ C_k such

that $C_i \neq \phi$ for i = 1, . . .K. (1)

$C_i \cap C_j = \phi$ for i, j = 1, . . .K; $i \neq j$ (2)

$$\Upsilon_{i=1}^{K} C_i = S \tag{3}$$

and the objects belonging into same cluster are similar in the sense of the given metric, while the objects belonging into different clusters are dissimilar in the same sense.

In other words, we seek a function $f : S \rightarrow \{1,\ldots\ldots,K\}$

such that for $i = 1,\ldots..,K : C_i = f^{-1}(i)$ where C_i satisfies the above conditions. The Euclidean metric can be used to measure the cluster quality.

Then the function f is sought such that

$$f \quad = \arg\min Evq(c_1,\ldots\ldots.,c_k) \tag{4}$$

$$= \arg\min_i \sum_{i=1}^{k} \| \vec{x}_i - \vec{c}_{f(\vec{x}_i)} \|^2$$

Where

$$\vec{c}_k = \frac{1}{|C_k|} \sum_{\vec{x} \in c_{k_i}} \vec{x}_i \qquad \text{k=1,…,K} \tag{5}$$

Therefore instead of function f directly, one can search for the centers of the clusters, i.e. vectors $\vec{c}_i,\ldots\ldots.,\vec{c}_k$.

implement the function f as

$$f(\vec{x}) = \arg\min_i \| \vec{x}_i - \vec{c}_i \|^2 \tag{6}$$

That is, assign the point to the cluster corresponding to the nearest centre.[21]

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. K-means algorithm

This algorithm is based on an adaptive approach where a random set of cluster base is selected from the original dataset, and each element update the nearest element of the base with average of its attributes.[7]

K-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. This algorithm aims at minimizing an objective function; in this case a squared error function.

The objective function.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \| x_i^{(j)} - c_j \|^2 \tag{7}$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the *n* data points from their respective cluster centers. The algorithm is composed of the following steps:

1. Place K points into the space represented by the objects that are being clustered. These points represent *initial group centroids.*
2. *Assign each object to the group that has the closest centroid.*
3. *When all objects have been assigned, recalculate the positions of the K centroids.*
4. *Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.*

B. Fuzzy C-means algorithm.

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method is frequently used in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m}\left\|x_i - c_j\right\|^2 \qquad (8)$$

$$1 \le m < \infty$$

where *m* is any real number greater than 1, *uij* is the degree of membership of *xi* in the cluster *j*, *xi* is the *i*th of d-dimensional measured data, *cj* is the d-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center. The algorithm is composed of the following steps:

1. *Initialize U=[uij] matrix, U(0)*
2. *At k-step: calculate the centers vectors C(k)=[cj] with U(k)*

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \qquad (9)$$

3. *Update U(k) , U(k+1)*

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\|x_i - c_j\right\|}{\left\|x_i - c_k\right\|}\right)^{\frac{2}{m-1}}} \qquad (10)$$

4. *If || U(k+1) - U(k)||< ε then STOP; otherwise return to step 2.*

IV. Proposed Algorithm : Self Adaptive GA With Clustering

Real Coded Self Adaptive Genetic Algorithm proposed in [4], was used for Multimodal Function Optimization. The algorithm has given good performance on Multimodal Functions. But fails to solve Multimodal Multipeak Functions. Here in this work, we propose Self Adaptive Genetic Algorithm which makes use of Clustering Algorithm.

Self Adaptation is important feature of Proposed Algorithm. MPX and MLX are two Multi-parent Crossover Operators. The MPX is exploitative and MLX is explorative in nature, a mechanism of utilizing suitable one is required to achieve efficient search. The basic idea of mechanism is that the probabilities of applying each operator are adapted based on the performance of the offspring generated by operator.[15]

In Self Adaptive GA with Clustering, K-Mean as well as Fuzzy C-means clustering algorithm are used. Performance of proposed Algorithm with K-means and with Fuzzy C-means is analyzed. Clustering is used after Population Initialization. The number of Clusters is a parameter of the Algorithm.

A. Algorithm: Self Adaptive GA with Clustering

In the terms of algorithm-generator, Self Adaptive GA with Clustering can be described as follows:

Step 1:- Generate Initial Population randomly.

Step 2:- Divide the Initial Population into Clusters by using Clustering Algorithm.(K-means or Fuzzy C-means.)

Step 3:-In each Cluster we have to apply the following steps.

Step 4:- The best parent and other (μ-1*)* random solutions are picked from *B* to form set P.

Step 5:- Create λ offspring solutions (the set C) from *P* using MPX or MLX operator. Choose recombination operator according to their probability. This probability depends upon the good offspring produced by the respective operator and is updated after every fixed number of function evaluations.

Step 6:- Replacement Plan: form a family of λ and μ solutions (the Set *R*).

Step 7:- Update Plan: From the *R*, choose *r* best solutions and put them in *R'*. Update *B* as

B := (*B*\R) U R'

Step 8:-Repeat the Steps 4 to 8 until Stopping Criteria met.

Step 9- Find best solution in each cluster.

Here set B denotes the population of solutions of size N. Operator adaptation methods is implemented as follows

1. Each member of population contains information about the operator (1 for MPX, 2 for MLX and 0 initially) along with decision variables and objectives. This information is placed into chromosomes at the time of generation.

2. In order to implement a ‘learning-rule’ to adapt operator probabilities two global counters; one for MPX operator and other for MLX operator are maintained. Increment these counters when offspring replaces parent. Operator is rewarded if offspring produced by it replaces best parent. Probability of operator is calculated as

$$P_{MPX} = (C_{MPX} / C_{MPX} + C_{MLX})$$
$$P_{MLX} = 1 - P_{MPX} \qquad (11)$$

Where P_{MPX} and P_{MLX} are probabilities of MPX and MLX respectively. C_{MPX} and C_{MLX} are global counters for MPX and MLX respectively. We limit the range of P_{MPX} to 0.05 to 0.95.

V. Experimental Setup

TABLE I. GA SETUP USED FOR EXPERIMENTATION

GA type	**Steady-state GA**
Population size (N)	150
Crossover probability parameter (p_c)	0.6 - 0.7 in step of 0.1
Distribution index (η)	For MPX=2 For MLX=4
Number of parents (*μ)*	10
Stopping criteria	Maximum 10^6 function evaluations or an objective value of Optima
Results average over	45 independent runs
Parameters for performance evaluation	1. Number of function evaluation for best run and worst run 2. Average number of function evaluation 3. Best fitness, Average fitness and Worst fitness 4. Number of runs converged to global Optima.
Number of children (λ)	2

When dealing with multi-modal functions, some modification is necessary to permit stable subpopulations at all peaks in the search space. So we have used two different techniques of Clustering K-means and Fuzzy C-means for dividing Population into subpopulations. Since we are comparing Self Adaptive GA with K-means Clustering and Self Adaptive GA with Fuzzy C-means Clustering it is mandatory to use the same setup for both the algorithms which is specified in table 1.

Optimization Functions f1, f2, f3, f4 and f5 are Minimization functions f6, f7 and f8 are Maximization functions.

VI. Results

We tested Self Adaptive GA with Clustering on some multimodal functions as listed in table 2. Results are shown in the table 3.Results obtained with RAGA in [4] have shown that, it has performed well for the Multimodal functions with single global peak i.e. functions f1 and f3. Algorithm searches global optima. But fails to solve Multimodal Multipeak Functions. Only single optima can be searched.

So clustering is introduced in the proposed Algorithm for forming subpopulations. Self Adaptive GA with KM have used K-means for Cluster formation. Each individual Cluster has evaluated to search the optima. Algorithm has given better performance while solving some multimodal multipeak functions with less number of Average Function evaluations. Figures 1 & 2 shows the Proper clusters formed by Self Adaptive GA with K-means. Each cluster contains a single optima. Figure 1(a) shows Initial Population, Figure 1(b) shows Clustered Population Figure 1(c) Optimas found for Five-Uneven-Peak Trap. Figure 2(a) shows Initial Population, Figure 2(b) shows Clustered Population Figure 2(c) Optimas found for Himmelblau’s function. But for some multimodal multipeak functions all optimas have not searched due to improper clustering as illustrated in figure 3(b).

To overcome this difficulty we have replaced K-means by Fuzzy C-means for Cluster formation. Self Adaptive GA with FCM has given better performance for all the functions listed in table 2 with less number of Average function evaluations. Using Fuzzy C-means, Proper clusters have been formed as illustrated in figures 3(c).

VII. Conclusion

In this paper, Self Adaptive GA with clustering is proposed. Algorithm employs both MPX and MLX operators. To choose crossover operator it has self-adaptive mechanism. It also uses replace worst parent with best offspring strategy in update plan. Multimodal Functions have multiple local as well as global optimas. When dealing with multi-modal functions, some modification is necessary to the standard GAs to permit stable subpopulations at all peaks in the search space. For this reason, we have used K-means or Fuzzy C-means Clustering algorithm for creating subpopulations.

Self Adaptive GA with K-means performed well for functions having less than five global optimas. But Self Adaptive GA with Fuzzy C-means performed better for all functions stated in table 2. The average function evaluations for Self Adaptive GA with Fuzzy C-means are less as compared with Self Adaptive GA with K-means. Comparative results are shown in table 3.

Future research will follow two parallel directions. The first one will extend the study of the current algorithm on real world problems. The second one will experiment on the application of other clustering algorithms, especially ones which do not require a priori knowledge of the number of clusters.

TABLE 2 MULTIMODAL TEST FUNCTIONS

Name	Test Function	Range	No. of Global Peaks
Michalewicz's function [16]	$f_1(x) = -\sum_{i=1}^{n} \sin(x_i) \cdot \left(\sin\left(\frac{i.x_i^2}{\Pi} \right) \right)^{2\cdot m}$	$0 \le x_i \le \prod$	1
Branins's rcos function [16]	$f_2(x_1, x_2) = a \cdot (x_2 - b \cdot x_1^2 + c \cdot x_1 - d)^2 + e \cdot (1 - f) \cdot \cos(x_1) + e$	$-5 \le x_1 \le 10$ $0 \le x_2 \le 15$	3
Goldstein-Price's function [16]	$f_3(x_1, x_2) = (1 + (x_1 + x_2 + 1)^2 \cdot (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2))$ $\cdot (30 + (2x_1 - 3x_2)^2 \cdot (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2))$	$-2 \le x_1x_2 \le 2$	1
Six-hump camel back function[16]	$f_4(x_1, x_2) = (4 - 2.1x_1^2 + x_1^{4/3}) \cdot x_1^2 + x_1x_2 + (-4 + 4x_2^2) \cdot x_2^2$	$-3 \le x_1 \le 3$ $-2 \le x_2 \le 2$	2
Himmelblau's function [20]	$f_5(x_1, x_2) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$	$-6 \le x_1x_2 \le 6$	4
Equal Maxima[20]	$f_6(x) = \sin^6(5\prod x)$	$0 \le x \le 1$	5
Uneven Maxima[20]	$f_7(x) = \sin^6(5\prod(x^{\frac{3}{4}} - 0.05))$	$0 \le x \le 1$	5
Five-Uneven-Peak Trap [20]	$f_8(x) = \begin{cases} 80(2.5 - x)\ for\, 0 \le x < 2.5 \\ 64(x - 2.5)\ for\, 2.5 \le x < 5.0 \\ 64(7.5 - x)\ for\, 5.0 \le x < 7.5 \\ 28(x - 7.5)\ for\, 7.5 \le x < 12.5 \\ 28(17.5 - x)\ for\, 12.5 \le x < 17.5 \\ 32(x - 17.5)\ for\, 17.5 \le x < 22.5 \\ 32(27.5 - x)\ for\, 22.5 \le x < 27.5 \\ 80(x - 27.5)\ for\, 27.5 \le x \le 30 \end{cases}$	$0 \le x \le 30$	2

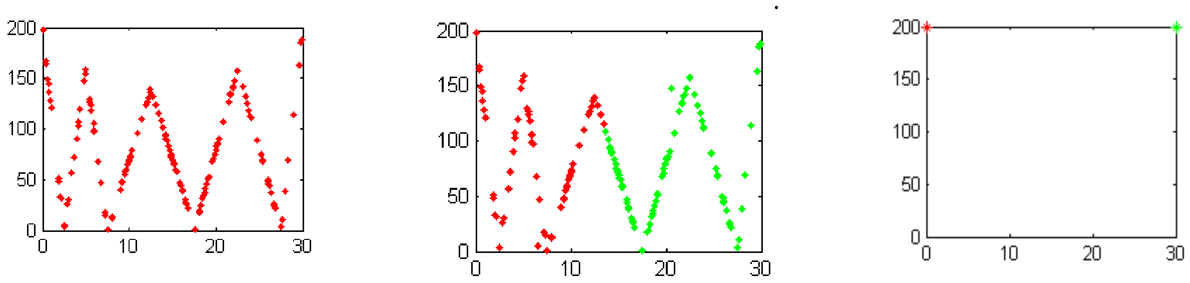


Figure 1.a) Initial Population b) Clustered Population using K-means c) Optima's found for Five-Uneven-Peak Trap

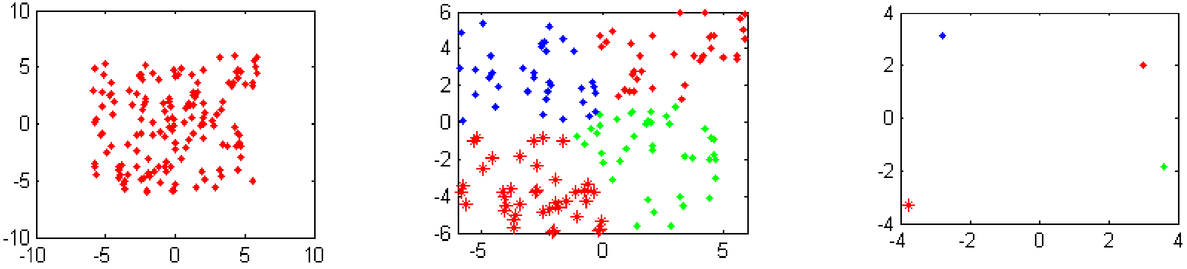


Figure 2.a) Initial Population b) Clustered Population using K-means c) Optima's found for Himmelblau's function

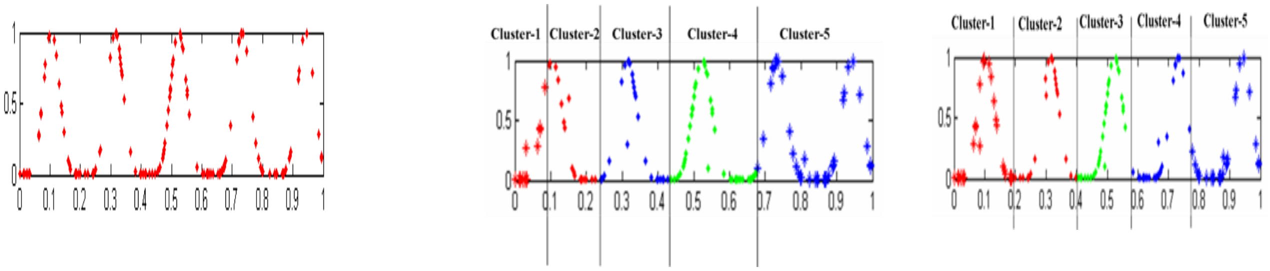


Figure 3.a) Initial Population b) Clustered Population using K-means c) Clustered Population using Fuzzy C-means for Equal Maxima Function.

TABLE 3 EXPERIMENTAL RESULTS OF RAGA WITH CLUSTERING

Sr.No.	Function Name	Algorithm	No. of Function Evaluations			Fitness			Success
			Best	Avg	Worst	Best	Avg	Worst	
1	f1 Michalewicz's function	RAGA	22226	107792	607670	-9.782	-9.68381	-9.66014	(45/45)
		SAGAKM	17044	157885	1000002	-9.76158	-9.68403	-9.6523	(41/45)
		SAGAFCM	20344	327997	1000002	-9.65959	-9.65091	-9.58876	(33/45)
2	F2 Branins's rcos function	RAGA	2094	3022.49	4358	0	0	0	(45/45)
		SAGAKM	560	1272.42	11048	0	0	0	(45/45)
		SAGAFCM	766	1224.82	2644	0	0	0	(45/45)
3	f3 Goldstein-Price's function	RAGA	2694	446178	1000002	3	34.2	84	(25/45)
		SAGAKM	2422	446169	1000002	3	31.8	84	(25/45)
		SAGAFCM	2590	446175	1000002	3	34.2	84	(25/45)
4	F4 Six-hump camel back function	RAGA	268	589.244	988	-1.03163	-1.03161	-1.0316	(45/45)
		SAGAKM	142	22598.4	1000002	-1.03163	-1.03162	-0.21546	(45/45)
		SAGAFCM	164	11528.8	1000002	-1.03163	-1.03162	-0.21546	(45/45)
5	f5 Himmelblau's function	RAGA	3110	3661.11	4250	8.72E-24	4.95E-21	9.89E-21	(45/45)
		SAGAKM	428	212338	1000002	1.94E-23	3.23E-21	16228.8	(45/45)
		SAGAFCM	734	1157.89	1874	4.59E-23	2.59E-21	9.99E-21	(45/45)
6	f6 Equal Maxima	RAGA	594	34285.9	1000002	1	1	1	(44/45)
		SAGAKM	78	10113	872798	1	1	1	(45/45)
		SAGAFCM	40	6327.58	964032	1	1	1	(45/45)
7	F7 Uneven Maxima	RAGA	212	710.133	1008	1	1	1	(45/45)
		SAGAKM	54	1263.51	101348	1	1	1	(45/45)
		SAGAFCM	30	345.511	10458	1	1	1	(45/45)
8	f8 Five-Uneven-Peak Trap	RAGA	4	5.95556	10	200	200	200	(45/45)
		SAGAKM	730	34537.9	1000002	200	200	160	(45/45)
		SAGAFCM	838	34520.2	1000002	200	200	160	(45/45)

REFERENCES

[1] Deb, K.: Multi-Objective Optimization using Evolutionary Algorithms, John Wiley & Sons, New York. (2001)

[2] Deb, K., Agrawal, S.: Understanding interactions among genetic algorithm parameters. Foundations of Genetic Algorithms V. Morgan Kaufmann(1999)

[3] Smith, J.E., Fogarty, T.C.: Operator and Parameter Adaptation in Genetic Algorithms. Soft Computing 1(2) (1997) 81–87

[4] M. M. Raghuwanshi and O. G. Kakde, “Real-Code Self-Adaptive Genetic Algorithm (RAGA)” Second Indian International Conference on Artificial Intelligence (IICAI-2005) 20-22 Dec. 2005 pp 3291-3302

[5] D. E. Goldberg, “Genetic Algorithms in Search, Optimization and Machine Learning”, Pearson Education Asia, 1989.

[6] M. M. Raghuwanshi and O. G. Kakde, “Multi-parent Recombination operator with Polynomial or Lognormal Distribution for Real Coded Genetic Algorithm” *2nd Indian International Conference on Artificial Intelligence (IICAI)*, pp. 3274-3290, 2005.

[7] Rahila Patel,M.M.Raghuwanshi,Anil N. Jaiswal “Modifying Genetic Algorithm with Species and Sexual Selection by using K-means Algorithm” 2009 IEEE International.Advance Comnputing Conference (LCC 2009) Patiala, India 6-7 March 2009

[8] Qing Lu, Changyong Liang, Enqiao Zhang “ Dynamic Sharing Scheme-based Multimodal Niche Genetic Algorithm” Proceedings of the 7th World Congress on Intelligent Control and Automation June 25 - 27, 2008, Chongqing, China

[9] Carlos M. Fonsecay and Peter J. Flemingz “Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization”

[10] Jian-Ping Li, Xiao-Dong Li and Alastair Wood “Species Based Evolutionary Algorithms for Multimodal Optimization: A Brief Review” WCCI 2010 IEEE World Congress on Computational Intelligence July, 18-23, 2010 - CCIB, Barcelona, Spain.

[11] Hartmut Pohlheim “GEATbx Introduction Evolutionary Algorithms: Overview, Methods and perators” (November 2005).

[12] Jingjun Zhang Yanmin Shang Ruizhen Gao Yuzhen Dong “An Improved Genetic Algorithm Based on K1 Triangulation with Variable Coefficient for Optimization of Multimodal Functions” ICIEA 2009

[13] Dr. K. Deb, Gulshan Singh: Comparison of multimodal optimization algorithms based on evolutionary algorithm. GECCO’06, Seattle, Washington USA.

[14] Deb, K., Agrawal, R.B.: Simulated binary crossover for continuous search space. Complex System 9 (1995) 115-148

[15] Raghuwanshi, M.M., Singru, P.M., Kale, U., Kakde, O.G.: Simulated Binary Crossover with Lognormal Distribution. In Proceedings of the 7th Asia-Pacific Conference on Complex Systems (Complex 2004)(2004)

[16] Hartmut Pohlheim: GEATbx: Example Functions (single and multi-objective functions) 2 Parametric Optimization.

[17] Rahila H. Sheikh, M. M.Raghuwanshi, Anil N. Jaiswal “Genetic Algorithm Based Clustering: A Survey” First International Conference on Emerging Trends in Engineering and Technology pp 314-319, 2008 IEEE

[18] Bruno Sareni and Laurent Kr¨ahenb¨uhl “Fitness Sharing and Niching Methods Revisited” IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, VOL. 2, NO. 3, SEPTEMBER 1998-97

[19] G. Harick. Finding multi-modal solutions using restricted tournament selection. In *Proceedings of the Sixth International Conference on Genetic Algorithms(ICGA-95)*, pages 24–31, 1997.

[20] Xiaodong Li, *Senior Member, IEEE* “Niching Without Niching Parameters: Particle Swarm Optimization Using a Ring Topology*”* IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, VOL 14, NO. 1, FEBRUARY 2010, pp 150-169

[21] Petra Kudov'a ,"Clustering Genetic Algorithm," IEEE, DOI 10.1 109/DEXA.2007.65, 2007

Neutrosophic Relational Database Decomposition

MeenaArora
CSE Department
JSS Academy of Technical Education
Noida, INDIA

Ranjit Biswas
Deptt. Of Faculty of Engg. & Technology
Manav Rachna University
Faridabad, INDIA

Dr. U.S.Pandey
Delhi University
Delhi,
INDIA

Abstract— **In this paper we present a method of decomposing a neutrosophic database relation with Neutrosophic attributes into basic relational form. Our objective is capable of manipulating incomplete as well as inconsistent information. Fuzzy relation or vague relation can only handle incomplete information. Authors are taking the Neutrosophic Relational database [8], [2] to show how imprecise data can be handled in relational schema.**

Keywords— Neutrosophic Logic; Ranking of intervals; β-value of an interval; Neutrosophic set; Neutrosophic Relation; Rank Neutrosophic 1NF.

I. Introduction

In the real world there are vaguely specified data values in many applications, such as sensor information, Robotics etc. Now take an example, when we ask the opinion of an expert about certain statement, he or she may say that the possibility that the statement is true is between 0.5 and 0.7, and the statement is false is between 0.2 and 0.4, and the degree that he or she is not sure is between 0.1 and 0.3. Here is another example, suppose there are 10 voters during a voting process. In time t_1, three vote "yes", two vote "no" and five are undecided, using neutrosophic notation, it can be expressed as x(0.3,0.5,0.2); in time t_2, three vote "yes", two vote "no", two give up and three are undecided, it then can be expressed as x(0.3,0.3,0.2). That is beyond the scope of the intuitionistic fuzzy set. So, the notion of neutrosophic set is more general and overcomes the fore mentioned issues. In neutrosophic set, indeterminacy is quantified explicitly and truth membership, indeterminacy-membership and falsity membership are independent. This assumption is very important in many applications such as information fusion in which we try to combine the data from different sensors. Neutrosophy was introduced by Smarandache [7] . Neutrosophic set is a powerful general formal framework which generalizes the concept of the classic set, fuzzy set, vague set etc.

The normalization process takes a relational Schema through a series of test to check up whether it satisfies a certain normal form. Consider an instance of a relation schema. In real life situation, the data available are not always precise or crisp, rather it can be in any form like it can be in natural language, any imprecise data or you can say Neutrosophic data .Consequently if, at least one data is Neutrosophic, the relation schema cannot be called to be in proper relational form. The quest to manage imprecision's is equal to major driving force in the database community is the Ultimate cause for many research areas: data mining, semi structured data, and schema matching, nearest neighbor. Processing probabilistic data is fundamentally more complex than other data models. Some previous approaches sidestepped complexity. For example, consider an attribute SALARY (in $) of a relation schema EMPLOYEE. If a tuple value for this attribute SALARY is precise viz. 5000, then it is a single atomic (Indivisible) value. But if a tuple value is Neutrosophic viz. "Approximately 5000", and then it cannot be called an atomic value. Due to the importance of the need for supporting uncertain data several researchers have addressed this problem. A wide body of work deals with fuzzy modeling of uncertain data [10]

In this paper a method to decompose such relational Schemas is suggested.

II. Prelimnaries

Relational data model was proposed by Ted Codd's pioneering paper [5]. Since then, relational database systems have been extensively studied and a lot of commercial relational database systems are currently available [3, 4, 6]. This data model usually takes care of only well-defined and unambiguous data. However, when we talk about the imprecise data or imperfect information , it will fail to answer. But our Lay users may or may not be aware of imprecision. In order to represent and manipulate various forms of incomplete information in relational databases, several extensions of the classical relational model have been proposed [13, 9, 11, 5, 14, 15]. The vague set and vague logic proposed by Gau and Behurer provide a requisite mathematical framework for dealing within complete and imprecise information.

Consequently, there is a genuine necessity for the different large size organizations, especially for the industries, companies having worldwide business, to develop such a system which should be able to answer the users queries posed in natural language, irrespective of the QLs and their grammar, without giving much botheration to the users. Most of these types of queries are not crisp in nature, and involve predicates with neutrosophic hedges (with concentration or dilation).

Thus, these types of queries are not strictly confined within the domains always. The corresponding predicates are not hard as in crisp predicates. Some predicates are soft and thus to answer a query a hard match is not always found from the databases by search, although the query is nice and very real, and should not be ignored or replaced according to the business policy of the industry. To deal with uncertainties in searching match for such queries, fuzzy logic and rather vague logic [8] and Neutrosophic logic by Smarandache [7] will be the appropriate tool.

In this section, Author presents some preliminaries on the theory of Neutrosophic Logic and Neutrosophic sets (NS) which will be required for the progress of this paper.

A. Neutrosophic Logic

In the *Neutrosophic Logic* (which is a generalization of fuzzy logic, especially of intuitionistic fuzzy logic) every logical variable *x* is described by an ordered triple $x = (T, I, F)$, where *T* is the degree of truth, *F* is the degree of falsehood, and *I* the degree of indeterminacy (or neutrality, i.e. neither true nor false, but vague, unknown, imprecise), with *T, I, F* standard or non-standard subsets of the non-standard unit interval $]^-0, 1^+[$. In addition, these values may vary over time, space, hidden parameters, etc.

There is a genuine necessity to develop such a system which should be able to answer the users queries posed in natural language, without giving much botheration to the users.

Let A and B be two neutrosophic sets. One can say, by language abuse, that any element neutrosophically belongs to any set, due to the percentages of truth/indeterminacy/falsity involved, which varies between 0 and 1 or even less than 0 or greater than 1.

Thus: x(50,20,30) belongs to A (which means, with a probability of 50% x is in A, with a probability of 30% x is not in A, and the rest is undecidable); or y(0,0,100) belongs to A (which normally means y is not for sure in A); or z(0,100,0) belongs to A (which means one does know absolutely nothing about z's affiliation with A).

More general, x ((20-30), (40-45) ∪[50-51], {20, 24, and 28}) belongs to the set A, which means:

- With a probability in between 20-30% x is in A (one cannot find an exact approximate because of various sources used);
- with a probability of 20% or 24% or 28% x is not in A ;
- the indeterminacy related to the appurtenance of x to A is in between 40-45% or between 50-51% (limits included);

The subsets representing the appurtenance, indeterminacy, and falsity may overlap, and n_sup = 30+51+28 > 100 in this case.

A logic in which each proposition is estimated to have the percentage of truth in a subset T, the percentage of indeterminacy in a subset I, and the percentage of falsity in a subset F, where T, I, F are defined below, is called *Neutrosophic Logic*. Constants: (T, I, F) truth-values, where T, I, F are standard or non-standard subsets of the nonstandard interval $]^-0, 1^+[$, where $n_{inf} = \inf T + \inf I + \inf F \geq {}^-0$, and $n_{sup} = \sup T + \sup I + \sup F \leq 3^+$. Neutrosophic logic [7] was created by Florentin Smarandache (1995)

B. Neutrosophic Sets

Neutrosophic set is a powerful general formal framework which generalizes the concept of the classic set, fuzzy set [1], Vague set [12] etc. A neutrosophic set A defined on universe U. x = x(T,I,F) ∈ A with T, I and F being the real standard or non-standard subsets of $]^-0,1^+[$, T is the degree of truth-membership of A, I is the degree of indeterminacy membership of A and F is the degree of falsity-membership of A.

Let X be a space of points (objects), with a generic element in X denoted by x. A neutrosophic set A in X is characterized by a truth-membership function T_A, an indeterminacy-membership function I_A and a falsity-membership function F_A. $T_A(x)$, $I_A(x)$ and $F_A(x)$ are real standard or non-standard subsets of$]^-0, 1^+[$. That is

$$T_A : X \rightarrow]^-0, 1^+[, \quad (1)$$
$$I_A : X \rightarrow]^-0, 1^+[, \quad (2)$$
$$F_A : X \rightarrow]^-0, 1^+[. \quad (3)$$

There is no restriction on the sum of $T_A(x)$, $I_A(x)$ and $F_A(x)$ so ${}^-0 \leq \sup T_A(x)+\sup I_A(x) + \sup F_A(x) \leq 3^+$.

Definition1. (Complement) The complement of a neutrosophic set A is denoted by c(A) and is defined by

$$T_{c(A)}(x) = \{1^+\} - T_A(x), \quad (4)$$
$$I_{c(A)}I(x) = \{1^+\} - I_A(x) , \quad (5)$$
$$F_{c(A)}(x) = \{1^+\} - F_A(x). \quad (6)$$

for all x in X.

Definition2. (Union) The union of two neutrosophic sets A and B is a neutrosophic set C, written as C = A ∪ B, whose truth-membership, indeterminacy-membership and falsity-membership functions are related to those of A and B by

$$T_C(x) = T_A(x) + T_B(x) - T_A(x) \times T_B(x), \quad (7)$$
$$I_C(x) = I_A(x) + I_B(x) - I_A(x) \times I_B(x), \quad (8)$$
$$F_C(x) = F_A(x) + F_B(x) - F_A(x) \times F_B(x). \quad (9)$$

for all x in X.

Definition3. (Intersection) The intersection of two neutrosophic sets A and B is a neutrosophic set C, written as C = A ∩ B, whose truth-membership, indeterminacy-membership and falsity-membership functions are related to those of A and B by

$$T_C(x) = T_A(x) \times T_B(x), \quad (10)$$
$$I_C(x) = I_A(x) \times I_B(x), \quad (11)$$
$$F_C(x) = F_A(x) \times F_B(x). \quad (12)$$

for all x in X.

C. Neutrosophic relation

In this section, we will define the Neutrosophic relation. A relation is basically set of attributes(columns) and tuples(records/rows) having atomic values i.e indivisible and consistent vales. A tuple in a neutrosophic relation is assigned a measure that will be referred to as the *truth* factor and also as the *false* factor. The interpretation of this measure is that we believe with confidence and doubt with confidence that the tuple is in the relation. The truth and false confidence factors for a tuple need not add to exactly 1. This allows for incompleteness and inconsistency to be represented.

III. CLASSICAL RELATIONAL MODEL APPROACH

A classical relational database [9] consists of a collection of relations. A relation is a table of values where each row represents a collection of related data values. In a table, each row is called a tuple, a column header is called an attribute and the table as a whole is called the relation. A relation schema R(A_1, A_2, …A_n) consists of a relation name R and list of attributes A_1, A_2, …A_n. There are various restrictions on data in the form of constraints. Domain constraints specify that each value of an A_i must be an atomic value from the domain dom(A_i)attribute. This includes restrictions on data types, on the range of values (if any), and on format of data.

IV. RANKING OF INTERVALS

Intervals are not ordered. Owing to this major weakness, there is no universal method of ranking a finite (or infinite) number of intervals. But in real life problems dealing with intervals we need to have some tactic to rank them in order to arrive at some conclusion. A method of ranking of intervals is used here in subsequent section.

Consider a decision maker (or any intelligent agent like a company manager, a factory supervisor, an intelligent robot, an intelligent network, etc.) who makes a pre-choice of a decision parameter $\beta \in [0, 1]$. The intervals are to be ranked once the decision-parameter β is fixed. But ranking may differ if the pre-choice β is renewed.

***Definition 4* (β-value of an interval)**
Let J = [a; b] be an interval. The β-value of the interval J is a non-negative real number J, given by $J_\beta = (1 - \beta).a + \beta.\ b$.

Clearly, $0 \le J_\beta \le 1$, and β= 0, $J_\beta = a$ which signifies that the decision-maker is pessimistic, and also for β= 1, $J_\beta = b$ which signifies that the decision maker is optimistic. For β= .5 it is the arithmetic-mean to be chosen usually for a moderate decision.

Comparison of two or more intervals we will do here on the basis of ¯-values of them. If the value of βis renewed, the comparison results may change. The following definition will make it clear.

***Definition 5*(Comparing two intervals)**
Let $J_1 = [a\ ,\ b]$*and* $J_2 = [c\ ,\ d]$ *be two intervals. Then for a chosen* $\beta \in [0\ ,1]$,*we define*

(i) $J_1 < J_2$, *if* $(J_1)_\beta < (J_2)_\beta$.
(ii) $J_1 > J_2$, *if* $(J_1)_\beta > (J_2)_\beta$.
(iii) $J_1 = J_2$, *if* $(J_1)_\beta = (J_2)_\beta$.

(Note: The intervals $J_1 = [a, b]$ and $J_2 = [a\ ,\ b]$ are strictly equal. For the other cases of the equality "$J_1 = J_2$", a further internal ranking could be done on the basis of their range i.e., interval-length. If range is more, we impose that the corresponding interval is greater).

So , the relational table where the ranking is involved , we name it by Rank Neutrosophic Table and the corresponding Rank Neutrosophic Normalisation is defined in the next section.

V. RANK NEUTROSOPHIC – 1NF OR (RNF)

In this section we have presented a method of decomposing a relational schema with Neutrosophic attributes into basic relational form. This Method is called as Rank Neutrosophic-First Normal Form -1NF(RNF) a revision of First normal Form in Relational database.

ALGORITHM

Let us present Sequence of steps for Rank Neutrosophic normalization of relation schema into 1NF(RN) :-

1) Remove all the Neutrosophic-attributes from the relation.

2) For each Neutrosophic-attribute create one separate table with the following attributes:

a) All attributes in the primary key

b) MV (z) (membership value)

c) NMV (Z) (non-membership value)

3) For every precise value of the Neutrosophic attribute put MV=1 and NMV=0.

Thus, if there is m number of attributes in the relation schema then, after normalization there will be in total (m + l) number of relations .In special case, when the hesitation or in deterministic parts are nil for every element of the universe of discourse the Neutrosophic number reduces to fuzzy number. In such cases, the attribute non membership value i.e. NMV (Z) will not be required in any reduced tables of 1NF.

The method of normalizing a relational schema (with Neutrosophic attributes) into 1NF is explained in this section. For the sake of simplicity, consider a relation schema R as given in Table I. with only one Neutrosophic attribute and all other three attributes being crisp. "Neutrosophic attribute" means that at least one attribute value in a relation instance is Neutrosophic.

TABLE I. RELATIONAL SCHEMA R

A_1	A_2	A_3	A_4

This relational schema R has four attributes of which say A4 is the only Neutrosophic attribute. Consider a relation instance r of R as shown in Table II. :

TABLE II. RELATIONAL TABLE R

A_1	A_2	A_3	A_4
A_{11}	A_{21}	A_{31}	A_{41}
A_{12}	A_{22}	A_{32}	A_{42}
A_{13}	A_{23}	A_{33}	A_{43}
A_{14}	A_{24}	A_{34}	A_{44}

Suppose that A2 is the primary key here, all the data are precise except $\tilde{a}$, which is an Neutrosophic number. Thus all the data except $\tilde{a}$ is atomic. This is not in 1NF because of the non atomic data $\tilde{a}$.

The neutrosophic number is the Neutrosophic set of the set R of real numbers. The universe of discourse R is an infinite set. But, in this method of normalization we shall consider a finite universe of discourse, say X, whose cardinality is N.

Let us suppose that X: $\{x_1, x_2,\ldots x_n\}$ and the Neutrosophic number $\tilde{a}$, proposed by a database expert is a NS (Neutrosophic Set) given by:

$\tilde{a} = \{ (x_i , \mu_i , V_i) : X_i \in X , I= 1 , 2 , 3 , . . .N\}.$
Then the Table II can be replaced by the following table III :

TABLE III. THE RELATIONAL INSTANCE R

A_1	A_2	A_3	A_4
A_{11}	A_{21}	A_{31}	A_{41}
A_{12}	A_{22}	A_3 2	$\{(Xi , \mu_i, Vi) , (X_1, \mu_1 ,V_1), \ldots (X_n, \mu_n , V_N)\}$
A_{13}	A_{23}	A_{33}	A_{43}
A_{14}	A_{24}	A_{34}	A_{44}

Now remove all the Neutrosophic attributes (here A_4 only), from Table III. Replace Table III by the following two tables Table IV. and Table V. :

TABLE IV. THE RELATION R_1

FNAME	**FID**	**YEARLY SALARY (in thousands of dollars)**
Juhi	F001	**4563**
Manoj	F002	**6789**
Geeta	F003	**Approximately 56**
Arun	**F004**	**8987**

TABLE V. THE RELATION R_2

A_2	A_4	MV (A4)	NMV (A4)
A_{21}	A_{41}	1	0
A_{22}	X_1	μ_1	V_1
A_{22}	X_2	μ_2	V_2
A_{22}	X_3	μ_3	V_3
….	…	….	…
A_{22}	X_n	μ_n	V_n
A_{23}	A_{43}	1	0
A_{24}	A_{44}	1	0

In Table V. we have all the attributes of the primary-key of r(here only one attribute A_2), the Neutrosophic attribute A_4 and two new attributes which are Membership _Value(A_4) or MV(A_4)and Non Membership _Value(A_4) or NMV(A_4). Corresponding to all precise values of A_4, the MV(A_4) value is put 1 and the NMV(A_4) value is 0.

Now we see that the relation schema is in 1NF. Such a method of normalization is called Neutrosophic normalization and the normal form is called Neutrosophic 1NF or 1NF (N).

VI. EXAMPLE WITH HYPOTHETICAL DATA

We study the method here by an example with hypothetical data.
Consider an example with hypothetical data. Consider a relation schema FACULTY as shown in Table VI. below whose primary key is FID and the attribute YEARLYSALARY is a Neutrosophic attribute.

TABLE VI. THE RELATION SCHEMA- FACULTY

FNAME	**FID**	**YEARLY SALARY (in thousands of dollars)**

Consider a relation instance of this relation schema given by the following Table VII.

TABLE VII. THE RELATIONAL TABLE - FACULTY

A_1	A_2	A_3
A_{11}	A_{21}	A_{31}
A_{12}	A_{22}	A_{32}
A_{13}	A_{23}	A_{33}
A_{14}	A_{24}	A_{34}

In this instance FNAME and FID are crisp attribute whereas YEARLY-SALARY is neutrosophic attribute, all the attribute values for FNAME are atomic; all the attribute values for the attribute FID are atomic. But all the attribute values for the attribute Yearly Salary are not atomic. The data "approximately 56" is an Neutrosophic number $5\tilde{6}$. Suppose that for this relation, a database expert proposes the Neutrosophic number $5\tilde{6}$ as an NS given by $5\tilde{6}$ **={(55,.8,.1),(56,.9,.03),(56.5,.7,.10)}.**

Therefore Table VII could be replaced by the following Table VIII.

TABLE VIII. RELATIONAL TABLE WITH NEUTROSOPHIC ATRIBUTES.

FNAME	**FID**	**YEARLY SALARY (in thousand of dollars)**
Juhi	F001	**4563**
Manoj	F002	**6789**
Geeta	F003	**{(**55,.8,.1),(56,.9,.03),(56.5,.7,.10)}
Arun	F004	**8987**

Now remove the Neutrosophic attribute YEARLY SALARY (YS) for this instance and divide it into two relations given as in Table IX and Table X.

TABLE IX. FACULTY-1 RELATION

FNAME	**FID**
Juhi	F001
Manoj	F002
Geeta	F003
Arun	F004

TABLE X. FACULTY-2 RELATION

FID	**YS**	**MV(YS)**	**NMV(YS)**
F001	**4563**	**1**	**0**
F002	**6789**	**1**	**0**
F003	**55**	**.8**	**.1**
F003	**56**	**.9**	**.03**
F003	**56.5**	**.7**	**.10**
F004	**8987**	**1**	**0**

Clearly, it is now in 1NF, called by 1NF (RN). For FACULTY-1, the Primary Key is FID, but for the newly created FACULTY-2 the Primary Key is {FID, YS}.

VII. CONCLUSION

In the above section we have presented a method of normalization of a relational schema with Neutrosophic attribute in 1NF (RN). In recent years neutrosophic algebraic structures have been investigated (see for instance [16], while the neutrosophic framework has found practical applications in a variety of different fields, such as relational database systems**.** We have implemented the method of normalization by an example given in section IV which proves that how the imprecise data can be handle in relational schema using First Normal Form of Rank Neutrosophic databases. Thereby we claim that the algorithm suggested in section V is totally a new concept which can easily handle the neutrosophic attributes of First normal Form.

REFERENCES

[1] Atanassov, K., Intuitionistic Fuzzy Sets : Theory and Applications, Physica-Verlag (2000), New- York.1955.

[2] Biswas,R ., Intuitionistic fuzzy relations Bull. Sous. Ens.Flous.Appl. (BUSEFA L) 70 ,1997,22-29.

[3] Chiang D., Chow L. R. and Hsien N,"Fuzzy information in extended fuzzy relational databases", Fuzzy Sets and Systems 92, pp.1-10.

[4] Codd, E.F., A relational model of data for large shared data banks, /Communications of the ACM, Vol.13 (6), p.377-387, June 1970 .

[5] E. F. Codd, Extending the Database Relational Model to Capture More Meaning, ACM Trans. Database Systems , 4(4):397-434,Dec. 1979.

[6] Elmasri and Navathe, Fundamentals of Database Systems, Addison-Wesley,New York,third edition,2000.

[7] F. Smarandache (2002a), A Unifying Field in Logics: Neutrosophic Logic, in Multiple-Valued Logic / An International Journal , Vol.8, No.3,385-438,2002.

[8] Gau, W.L. and Buehrer, D, J., Vague sets, IEEE Transactions on Systems, Man and Cybernetics, Vol.23 (1993) 610-614

[9] J. Biskup, A Foundation of Codd'sRelationalMaybe-operations, ACM Trans. Database Systems, 8, 4:608-636, Dec. 1983.

[10] J. Galindo, A. Urrutia, and M. Piattini, "Fuzzy Databases: Modeling, Design, and Implementation". Idea Group Publishing, 2006.

[11] M. L. Brodie, J. Mylopoulous, and J. W.Schmidt, On the Development of Data Models, On Conceptual Modeling , 19-47, 1984

[12] Meena Arora and Ranjit Biswas, Deployment of Neutrosophic technology to retrieve answers for queries posed in natural language, in 3rd International Conference on Computer Science and Information Technology ICCSIT 2010, IEEE catalog Number CFP1057E-art, volume No. 3, ISBN: 978-1-4244-5540-9, Pages : 435-439, 2010.

[13] S. Parsons, Current Approaches to Handing Imperfect Information in Data and Knowledge Bases, IEEE Trans, Knowledge and Data Engineering , 3:353- 372, 1996.

[14] W. Lipski, On Semantic Issues Connected with Incomplete Information Databases, ACM Trans. Database Systems, 4, 3:262-296, Sept.1979.

[15] W. Lipski, On Databases with Incomplete Information, Journal of the Association for Computing Machinery , 28:41-70, 1981.

W.B. Kandasamy, Smarandache Neutrosophic Algebraic Structures, Hexis, Phoenix, 2006.

[16] W.B. Kandasamy, Smarandache Neutrosophic Algebraic Structures, Hexis, Phoenix, 2006.

A rule-based Afan Oromo Grammar Checker

Debela Tesfaye
Information Technology,Jimma Institute of Technology
Jimma, Ethiopia

Abstract—**Natural language processing (NLP) is a subfield of computer science, with strong connections to artificial intelligence. One area of NLP is concerned with creating proofing systems, such as grammar checker. Grammar checker determines the syntactical correctness of a sentence which is mostly used in word processors and compilers.**
For languages, such as Afan Oromo, advanced tools have been lacking and are still in the early stages. In this paper a rule based grammar checker is presented. The rule base is entirely developed and dependent on the morphology of the language . The checker is evaluated and shown a promising result.

Keywords- afan oromo grammar checker; rule based grammar checker.

I. INTRODUCTION

Natural language processing (NLP) is a subfield of computer science, with strong connections to artificial intelligence. Natural language processing (NLP) is normally used to describe the function of computer system which analyze or synthesize spoken or written language [3]. One area of NLP is concerned with creating proofing systems, such as spell checkers and grammar checkers. A grammar checker looks for grammatical errors and, in many cases, suggests possible corrections. Grammar checking is one of the most widely used tools within language engineering. Spelling, grammar and style checking for English has been an integrated part of common word processors for some years now.

The great challenge of intelligent automatic text processing is to use unrestricted natural language to exchange information with a creature of a totally different nature: the computer. People now want assistance not only in mechanical, but also in intellectual efforts. They would like the machine to read an unprepared text, to test it for correctness, to execute the instructions contained in the text, or even to comprehend it well enough to produce a reasonable response based on its meaning. Human beings want to keep for themselves only the final decisions.

Millions and millions of persons dealing with texts throughout the world do not have enough knowledge and education, or just time and a wish, to meet the modern standards of document processing [4]. For example, a secretary in an office cannot take into consideration each time the hundreds of various rules necessary to write down a good business letter to another company, especially when he or she is not writing in his or her native language. It is just cheaper to teach the machine once to do this work, rather than repeatedly teach every new generation of computer users to do it by themselves.

Grammar checker determines the syntactical correctness of a sentence. Grammar checking is mostly used in word processors and compilers. Grammar checking for application like compiler is easier to implement because the vocabulary is finite for programming languages but for a natural language it is challenging because of infinite vocabulary.

A lot of work has gone into developing sophisticated systems that have gone into widespread use, such as automatic translators and spell checkers. However, most such programs are strictly commercial, and therefore there exists no documentation of the algorithms and rules used. For languages, such as Afan Oromo, advanced tools have been lacking and are still in the early stages. However, one of the most widely used grammar checkers for English, Microsoft Office Suite grammar checker, is also not above controversy [1]. It demonstrates that work on grammar checker in real time is not very easy task; so starting the implementation for language like Afan Oromo is a major feat. In this research, a rule based grammar checker for Afan Oromo is presented.

II. BACKGROUND

Detection and correction of grammatical errors by taking into account adjacent words in the sentence or even the whole sentence are much more difficult tasks for computational linguists and software developers than just checking orthography. Grammar errors are those violating, for example, the syntactic laws or the laws related to the structure of a sentence. In Afan Oromo, one of these laws is the agreement between a noun and an adjective in gender and grammatical number[7]. For example, in *Jarri dhufaa jira*, subject and verb disagree in number. Jarri(*they) which is the subject of the sentence is* plural, and the verb of the sentence *jira* is the indicator for third person singular masculine.

Three methods are widely used for grammar checking in a language; syntax-based checking, statistics-based checking and rule-based checking.

A. Syntax based checking

In this approach, a text is completely parsed, i.e. the sentences are analyzed and each sentence is assigned a tree structure. The text is considered incorrect if the parsing does not succeed.

B. *Statistics based checking*

In the statistical approach the system is trained on a corpus to learn what is 'correct'. In this approach, a POS-annotated corpus is used to build a list of POS tag sequences. Some sequences will be very common (for example determiner, adjective, noun as in the old man), others will probably not occur at all (for example determiner, determiner, adjective). Sequences which occur often in the corpus can be considered correct, whereas uncommon sequences might be considered as errors. This method has a few disadvantages. One of these is that it can be difficult to understand the error given by the system as there is not a specific error message. This also makes it more difficult to realize when a false positive is given [1].

C. *Rule based checking*

Using the rule-based approach to grammar checking involves manually constructing error detection rules for the language. These rules are then used to find errors in text that has already been analyzed, i.e. Tagged with a part-of-speech tagger. These rules often contain suggestions on how to correct the error found in the text.

A lot of work has gone into developing grammar checkers for different languages. The most progress, by far, has been made for English. The earliest grammar checkers for English were developed in the 1970s and have gradually been improving over the last decades. Although there is still room for improvement their use is quite widespread as an English grammar checker is built into the most used word processor today, Microsoft Word.

EasyEnglish is a grammar checker developed at IBM especially for non-native speakers. It is based on the English Slot Grammar. It finds errors by "exploring the parse tree expressed as a network" [5]. The errors seem to be formalized as patterns that match the parse tree. Unfortunately [5] does not explain what exactly happens if a sentence cannot be parsed and thus no complete tree can be built.

III. Overview of Afan Oromo

Afan Oromo is one of the major languages that is widely spoken and used in Ethiopia. Currently it is an official language of Oromia state (which is the largest region in Ethiopia). It is used by Oromo people, who are the largest ethnic group in Ethiopia, which amounts to 34.5% of the total population [6]. The language has become the official language in Oromia regional offices *a*nd is also instructional language starting from elementary to university level.

Like a number of other African and Ethiopian languages, Afan Oromo has a very rich morphology [7]. It has the basic features of agglutinative languages where all bound forms (morphemes) are affixes. In agglutinative languages like Afan Oromo most of the grammatical information is conveyed through affixes and other structures.

Therefore the grammatical information of the language is described in relation to its morphology. This makes it very hard to create grammar checker and develop general understanding of the language.

As any other language the grammar of Afan Oromo exhibits gender,number,cases,tenses etc. But the grammatical presentation of the above cases are different from other languages and exhibits its own structure. Unlike English, Afan Oromo gender,number,tense and other cases are described using affix. Therefore the grammatical rule is mostly dependent on the affixation rule of the language.

Example:

1. Inni kalleessa dhufe.(he came yesterday)
2. Isheen kalleessa dhufte.(she came yesterday)
3. Inni dhufaa jira.(he is coming)
4. Isheen dhufaa jirti.(she is coming)

In the above four sentences the gender and tense of the sentences are described through suffix which is attached to the verbs dhuf- and jir-.

IV. Afan Oromo grammar checker

As described above, Afan Oromo exhibits its own grammatical structure. Therefore it is not possible to apply and use the grammatical rule of another language for Afan Oromo grammar checker. In this paper different 123 rules were constructed and used in order to identify grammatical error of the language. With the use of these carefully constructed error detection rules, the system can detect and suggest corrections for a number of grammatical errors in Afan Oromo texts. Afan Oromo Grammar Checker has the following components:

A. *Tokenizer Module*

The tokenizer module splits the input text (paragraphs) from an input file into sentences. The tokenized sentences are further tokenized into words.

B. *Parts of Speech (POS) Tagger Module*

Part of speech taggers are very important for our approach. In POS tagging of a text, each word in the text is assigned a part of speech. We have used tagger based on Hidden Markov Model which uses a manually tagged corpus for training"unpublished"[8].

C. *Stammer Module*

A stemming algorithm is a procedure that reduces all words with the same stem to a common form by stripping of its derivational and inflectional suffixes. The stammer module of this checker provides the root and affix for the tagged words. Like a number of other African and Ethiopian languages, Afan Oromo has a very rich morphology [7]. It has the basic features of agglutinative languages where all bound forms (morphemes) are affixes. In agglutinative languages like Afan Oromo most of the grammatical information is conveyed through affixes (prefixes, infixes and suffixes) attached to the roots or stems. Both Afan Oromo nouns and adjectives are highly inflected for number and gender. In contrast to the English plural marker s (-es), there are more than 12 major and very common plural markers in Afan Oromo nouns (example: -oota, -ooli, -wwan, -lee, -an, -een, -oo, etc.) [2]. Afaan Oromo verbs are also highly inflected for gender, person, number and tenses.

The Afan Ormo stammer is based on a series of steps that each removes a certain type of affix byway of substitution rules. These rules only apply when certain conditions hold, e.g. the resulting stem must have a certain minimal length. Most rules have a condition based on the so-called measure. The measure is the number of vowel-consonant sequences (where consecutive vowels or consonants are counted as one) which are present in the resulting stem. This condition must prevent that letters which look like a suffix but are just part of the stem will be removed.

The affix-rules have the following general form:

Affix → substitution measure-condition <additional conditions>

Where:

Affix is a valid Afan Oromo prefix or suffix

In Afan Oromo repetition (plural) is formed by duplicating the first syllabus and it is also considered as prefix.

Substitution is a string which is substituted with a given affix to produce valid stem.

Measure-condition is the number of vowel-consonant sequences (where consecutive vowels or consonants are counted as one) which are present in the resulting stem.

Additional conditions- additional conditions are also designed to cover some specific phenomena. Examples of these conditions are, Endswith Vowel/Consonant.

D. *Grammatical Relation Finder*

Assigns grammatical relations between subject and verb,subject and adjective,main verb and subordinate verb in terms of number,gender and tense. In this paper 123 different rules are constructed and presented. This rules takes the affixes that are identified and separated from a root word using the stammer module in order to identify the agreement between subject and verb, subject and adjective, main verb and subordinate verb in number, tense, gender and other causes. As explained in section III the grammatical information of the above cases are presented using affix-rules in the language.

E. *Suggestions creating module*

It provides the correct sentence alternatives. This module provides the alternatives in two way directions. For example in the case of subject verb disagreement it provides one or more alternatives by adjusting the subject and leaving the verbs as they are. Or provide one or more alternatives by adjusting the verbs and leaving the subject as it is. This is based on the assumption that errors can be committed both on the subject and the verb. There for the users must be provided with correct alternative by correcting either of the two one at a time.

The Grammar Checker General Algorithm

The algorithm has five steps as presented in the following section:

step 1.

Tokenize the sentence using '.' or '!' or '?'

step 2.

Identify the part of the speech of each token in the sentence

Step 3:

Identify the root and affixes of each token(word)

Step 4:

4: a:

Forward the affixes to the rule based that checks Subject-verb,subject and adjective,main verb and subordinate verb agreement in terms of number,tense and gender agreement.

4:b:

Check for punctuation errors.

Step 5:

provide grammatically correct sentence suggestions.

The algorithm is illustrated using the following example.

Inni saree ajjeste. Because of grammatical error the sentence is meaningless.

Step 1:

The tokenizer module identify the tokens as :

Inni, saree, and ajjeste.

Step 2:

The part-of-speech tagger tagged the tokens as:

Inni as the subject of the sentence, Saree as the object and ajjeste as the verb.

step 3:

The stammer module identified the root and affix of the word ajjeste as:

ajjes- is the root and -te is the suffix.

Inni has no any suffix.

Saree is the object of the sentence. In Afan Oromo object of the sentence has no any grammatical relation with the subject and verb of the sentence.

step 4:

The subject of the sentence is 1st person singular masculine and the suffix is feminine marker.

The rule

If the subject of the sentence is 3st person singular masculine (Inni) then the verb must end with the masculine marker suffixes –a, and -e.

So the rule-based marks for the subject-verb verb disagreement.

Step 5:

T*he correct suggestions are:*

a. Inni saree ajjese. By changing the suffix from feminine marker to masculine marker e.

b. Isheen saree ajjeste. By changing the subject of the sentence from masculine Inni to feminine Isheen.

Sample rules of the grammar checker

Definitions:

sg.1.p=first person singular

2.p = second person

3.p.m= third person masculine

3.p.f=third person feminine

2..p.pl=second person plural

3.p.pl=third person plural

RV=root verb

There are a total of 123 rules. The rules ranging from 81 to 86 covers past perfect tense as presented in the following.

rule 81.	*sg. l.p +RV+een+ture*
rule 82.	*2.p.sg + RV+tee+turte*
rule 83.	*3.p.m.+RV+ee+ture*
RULE 84.	*3.p.f.+RV+tee,dee+turte*
RULE 85.	*2.p.pl +nee++turre*
RULE 86.	*3.p.pl+ani++turan*

Explanation of the rules:

If the subject of the sentence is first person singular, the verb(s) must end with the suffix -een and the sentence must end with ture.

If the subject of the sentence is third person singular feminine, the verb(s) must end with the suffix -tee,dee and the sentence must end with turte.

Example: Callise bira dabruu yaal<u>*ee ture*</u>*.(he was trying to pass by silent). In the above example, the subject of the sentence is third person singular masculine, so the verb must end with the suffix -ee and the sentence must end with ture.*

V. EVALUATION OF THE CHECKER

Grammar and style checking software have involved measuring the program's error detection capacity in terms of precision (i.e. error detection correctness) and recall (i.e. error coverage) [9]; [10];[11].

In order to check the performance of the system a student graduation thesis text is used. A thesis work of Afan Oromo 1st degree graduate of 2011 were used in order to measure the performance of the checker. Originally, it was thought best to get some sort of text from non-native Afan Oromo speakers as it was assumed that students learning the language might not have the same 'feel for the language' as native speakers and therefore have more grammatical errors in their texts.

Finally, the above data were run through the grammar checker for errors. In order to calculate the performance rate the number of errors in the texts, number of errors found by the Grammar checker and the number of false positives generated by the grammar checker were counted. These numbers were then used to calculate the precision and recall of the system.

The table below shows the precision and recall rates for all the errors in the texts as well as the corresponding rates for each type of error. The rates are found as follows:

$$\text{Precision} = \frac{\text{number of correctly flagged error}}{\text{total number of flagged error}}$$

$$\text{Recall} = \frac{\text{number of correctly flagged error}}{\text{total number of error that occur in the text}}$$

TABLE I. PERFORMANCE RESULT

	Measuring criterias					
	Incorrect flags	*Correct flags*	*Total Number of flags*	Total number of errors in the test set	precision	recall
In NO/ %	100	800	900	1000	88.89%	80.00 %

There are several reasons why a false alarm might occur:

- The stammer identified the root and affix of some words incorrectly.
- A word has been assigned an incorrect part-of-speech tag.
- The rule is not complete and didn't covered every case.

VI. CONCLUSION

In this thesis, Afan Oromo grammar checker has been developed and tested on real-world errors. As can be seen from table 1.1 the performance of the checker is promising. Most of the false flags are related to compound, complex and compound complex sentences as most of the rules are constructed for simple sentences. More rules that handles the above listed types of sentences can be added to the existing rules in order to improve the performance of the grammar checker. There are also sentences that exhibits grammatical errors but not flagged by the checker.

Other than the incompleteness of the rules the part-of-speech tagger component of the checker has also provided incorrectly tagged words for the checker. The incorrectly tagged words lead the checker to not flag errors and generate false flags. The stammer component that separates the root from affix of a word is important since the grammar of the language is mostly described through affixes. Generally, since the grammar rules in the grammar checker is largely dependent on morphology of the language this approach is believed too be used for other languages that are rich in morphology.

REFERENCES

[1] Naber, D. (2003). A Rule-Based Style and Grammar Checker. Diplomarbeit. Technische Fakultät Bielefeld.

[2] Debela Tesfaye & Ermias Abebe(2010). Designing a rule based stemmar for afan Oromo text. International Journal of Computational Linguistics (IJCL), Volume (1): Issue (2)

[3] Peter Jackson and Isabelle Moulinier, natural language processing for online applications: text retrieval, extraction and categorization,1984

[4] Igor Balshagov and Alexander Gelbukh, computational linguistics models resources and applications,2004

[5] Arendse Bernth: EasyEnglish: Grammar Checking for Non-Native Speakers, Proceedings of the Third International Workshop on Controlled Language Applications (CLAW00), Association for Computational Linguistics, April 29-30, Seattle, Washington, pp. 33-42, 2000

[6] Census Report. "Ethiopia's population now 76 million", http://ethiopolitics.com/news, (2008)

[7] Gumii Qormaata Afaan Oromoo. "Caasluga Afaan Oromoo Jildi I", Komishinii Aadaaf Turizmii Oromiyaa, Finfinnee, Ethiopia, pp. 105-220 (1995)

[8] C. G. Mewis. A Grammatical sketch of Written Oromo, Germany: Koln,pp. 25-99 (2001)

[9] Getachew Mamo. Statistical model Part-of-speech Tager for afan Oromo,Addis Abeba University,2009.

[10] Kukich, K. Techniques for automatically correcting words in text. ACM Computing surveys, Vol. 24, No. 4, pp. 377–439 ,1992.

[11] Birn, J. Detecting grammar errors with Lingsoft's Swedish grammar checker. In Proc. 12th Nordic Conference in Computational Linguistics, Nodalida-99. Trondheim, pp. 28–40, 2000.

[12] Richardson, S & Braden-Harder, L. The Experience of Developing a Large-Scale Natural Language Processing System: Critique. In Jensen, K. Heidorn, G. E. Richardson, S. D. (eds.), Natural Language Processing: The PLNLP Approach, pp. 77- 89, 1993.