

ISSN: 2250-3005 (online version)

 IJCER

ONLINE PEER REVIEWED JOURNAL

International Journal of Computational
Engineering Research

International Journal
of Computational Engineering Research

International Journal of Computational Engineering Research (IJCER)



Published by:
IJCER

Volume 3, Issue 8, August,

Editorial Board

Editor-In-Chief

Prof. Chetan Sharma

Specialization: Electronics Engineering, India
Qualification: Ph.d, Nanotechnology, IIT Delhi, India

Editorial Committees

DR.Qais Faryadi

Qualification: PhD Computer Science
Affiliation: USIM(Islamic Science University of Malaysia)

Dr. Lingyan Cao

Qualification: Ph.D. Applied Mathematics in Finance
Affiliation: University of Maryland College Park,MD, US

Dr. A.V.L.N.S.H. HARIHARAN

Qualification: Phd Chemistry
Affiliation: GITAM UNIVERSITY, VISAKHAPATNAM, India

DR. MD. MUSTAFIZUR RAHMAN

Qualification: Phd Mechanical and Materials Engineering
Affiliation: University Kebangsaan Malaysia (UKM)

Dr. S. Morteza Bayareh

Qualificatio: Phd Mechanical Engineering, IUT
Affiliation: Islamic Azad University, Lamerd Branch
Daneshjoo Square, Lamerd, Fars, Iran

Dr. Zahéra Mekkioui

Qualification: Phd Electronics
Affiliation: University of Tlemcen, Algeria

Dr. Yilun Shang

Qualification: Postdoctoral Fellow Computer Science
Affiliation: University of Texas at San Antonio, TX 78249

Lugen M.Zake Sheet

Qualification: Phd, Department of Mathematics
Affiliation: University of Mosul, Iraq

Mohamed Abdellatif

Qualification: PhD Intelligence Technology
Affiliation: Graduate School of Natural Science and Technology

Meisam Mahdavi

Qualification: Phd Electrical and Computer Engineering

Affiliation: University of Tehran, North Kargar st. (across the ninth lane), Tehran, Iran

Dr. Ahmed Nabih Zaki Rashed

Qualification: Ph. D Electronic Engineering

Affiliation: Menoufia University, Egypt

Dr. José M. Merigó Lindahl

Qualification: Phd Business Administration

Affiliation: Department of Business Administration, University of Barcelona, Spain

Dr. Mohamed Shokry Nayle

Qualification: Phd, Engineering

Affiliation: faculty of engineering Tanta University Egypt

CONTENTS :

S.No	Title Name	Page No.
Version I		
1.	Numerical Study of the Capability of Various Turbulence Models to Predict the Heat Transfer Characteristics of Supercritical Water Flow P J Saji, S Suresh, R Dhanuskodi, P Mallikarjuna Rao, D Santhosh Kumar	01-07
2.	Moments of the Reliability $R = P(Y < X)$ As A Random Variable Mohammed A. Shayib , Aliakbar Montazer Haghighi	08-17
3.	The Integrated Locomotive Assignment and Crew Scheduling Problem Ash Aksoy, Alper Altan	18-24
4.	Search Engine Using Spatial Data P.Sreedevi, G.Sridevi, B.Padmaja	25-37
5.	Second Derivative Free Modification with a Parameter For Chebyshev's Method J. Jayakumar, P. Jayasilan	38-42
6.	Protecting Frequent Item sets Disclosure in Data Sets and Preserving Item Sets Mining Shaik Mahammad Rafi., M.Suman , Majjari Sudhakar, P.Ramesh , P.Venkata Ramanaiah	43-50
7.	Effect of Controlling Parameters on Heat Transfer during Jet Array Impingement Cooling of a Hot Steel Plate Purna C. Mishra, Santosh K. Nayak, Durga P. Ghosh, Manoj Ukamanal, Swarup K. Nayak, Susant K. Sahu	51-57
8.	Design and Testing Of Prefix Adder for High Speed Application by Using Verilog HDL Rajitha Chandragiri, P. Venkata Lavanya	58-62
9.	Risk Assessment of Bot Projects Sharmila Mane, Dr. S.S. Pimplikar	63-69
Version II		
1.	Image Fusion For Medical Image Retrieval Deepali Sale, Dr. Madhuri Joshi, Varsha Patil, Pallavi Sonare ,Chaya Jadhav	01-09
2.	Content Based Image Retrieval in Medical Imaging Prachi.Bhende, Prof.A.N.Cheeran	10-15
3.	Indian Two Wheeler Auto Industry and Concurrent Engineering Dr. V.Venkata Ramana , V.Visweswara Reddy, T.Govardhan Reddy	16-19

4.	Intrusion Detection In Manets Using Secure Leader Election Yasmeen Shaikh, V. K. Parvati	20-29
5.	Involute Gear Tooth Contact And Bending Stress Analysis Vishwjeet V.Ambade, Prof. Dr. A.V.Vanalkar, Prof.P.R.Gajbhiye	30-36
6.	Model based numerical state feedback control of jet impingement cooling of a steel plate by pole placement technique Purna Chandra Mishra, Swarup Kumar Nayak, Santosh K. Nayak, Premananda Pradhan, Durga P. Ghosh	37-44
7.	Remote Control System through Mobile and DTMF Abdiweli Abdillahi Soufi, Abdirasoul Jabar Alzubaidi	45-52
8.	Seismic Response of Multi-Story Structure With Semi-Active Multiple Tuned Mass Friction Damper Alka Y. Pisal, R. S. Jangid	53-72
9.	Overview Applications of Data Mining In Health Care: The Case Study of Arusha Region Salim Diwani, ,Suzan Mishol ,Daniel S.Kayange ,Dina Machuve ,Anael Sam	73-77
10.	FPGA Implementation of DA Algritm for Fir Filter Putta.Solman Raju , Garigipati.Vijay Kumar , Yrk Paramahamsa	78-83
11.	A REVIEW ON REPLICA NODE DETECTION LGORITHMS IN WSNS C. Geetha	84-89

Numerical Study of the Capability of Various Turbulence Models to Predict the Heat Transfer Characteristics of Supercritical Water Flow

P J Saji¹, S Suresh², R Dhanuskodi³, P Mallikarjuna Rao⁴, D Santhosh Kumar⁵

^{1,2,4}Department of Mechanical Engineering, National Institute of Technology,
Tiruchirappalli-620015, India.

^{3,5}Research and Development Department, Bharat Heavy Electricals Limited,
Tiruchirappalli-600014, India.

ABSTRACT

A two-dimensional model is developed to numerically study the capability of various turbulence models to predict the heat transfer characteristics of supercritical water flowing through a vertical tube. At the tube inlet, mass flux of $1260 \text{ kg/m}^2\text{s}$ is applied. At the tube outlet, normal gradients of velocity, Turbulence Kinetic energy and Turbulence dissipation rate (k and ϵ) are assigned zero and a pressure of 245.2 bar is specified. Uniform heat flux condition ranging from 233 to 698 kW/m^2 is applied around the tube wall. After performing a grid independence test, a non-uniform mesh of 200 nodes along the radial direction and 600 uniform nodes along the axial direction was chosen. To capture the large variations of flow field variables near the tube wall, a fine grid is used close to the wall (wall $y^+ < 0.3$). A set of standard two equation models (Standard $k-\epsilon$, RNG $k-\epsilon$, $k-\omega$ SST) and 5 Low Reynolds Number (Low Re) models are chosen in the present study. The results are compared with the experimental results of Yamagata *et al.* (1972). The tube wall temperature was plotted as a function of the fluid enthalpy. It is seen that under moderate heat flux conditions, the average difference between the wall temperature prediction by various models and the experimental results are within $3 \text{ }^\circ\text{C}$. Standard $k-\epsilon$, RNG $k-\epsilon$ and Two Low Re models came closest to the experimental results under the given conditions. Considering the computational time taken, RNG $k-\epsilon$ model has been picked as the best choice. The model is also validated for its capability to work at higher heat fluxes.

KEYWORDS: Supercritical water, Turbulence modeling, Tube wall temperature prediction, Heat transfer characteristics

NOMENCLATURE

G	mass flux [$\text{kg}/(\text{m}^2\text{s})$]
k	turbulence kinetic energy [m^2/s^2]
q	heat flux [kW/m^2]
T	temperature [$^\circ\text{C}$]
u	velocity component [m/s]
p	pressure [bar]
y^+	dimensionless wall function
ϵ	turbulence dissipation rate [m^2/s^2]
τ	turbulence stress [N/m^2]

ω	Inverse turbulence time scale
<i>Abbreviations</i>	
RANS	Reynolds averaged Navier-Stokes
SST	Shear Stress Transport
SIMPLE	Semi-implicit method for pressure linked equations
QUICK	Quadratic upstream interpolation of convective kinematics

I. INTRODUCTION

Due to the acute variation in the thermo-physical properties of supercritical fluids (Figure 1) near their pseudo critical point, they are being used in a wide range of engineering applications such as thermal power generation, nuclear power generation, refrigeration applications, electronic applications, chemical applications, pharmaceutical applications etc. Most of the applications make use of the higher specific heat near the pseudo critical region. Even though research in supercritical applications started as early as 1930, the idea of using the same in thermal power generation was unleashed in the 1950's. 50's saw a series of pilot scale plants being set up in US and USSR. Improvement in overall unit efficiency at higher pressures was the driving point behind this trend. At higher turbine inlet pressure, the net mechanical energy available from the Rankine cycle increases compared to that from subcritical pressure. This is due to the improvement in the quality of energy at higher pressures, in other words it is due to the decrease in irreversibilities associated with energy conversion at higher pressures.

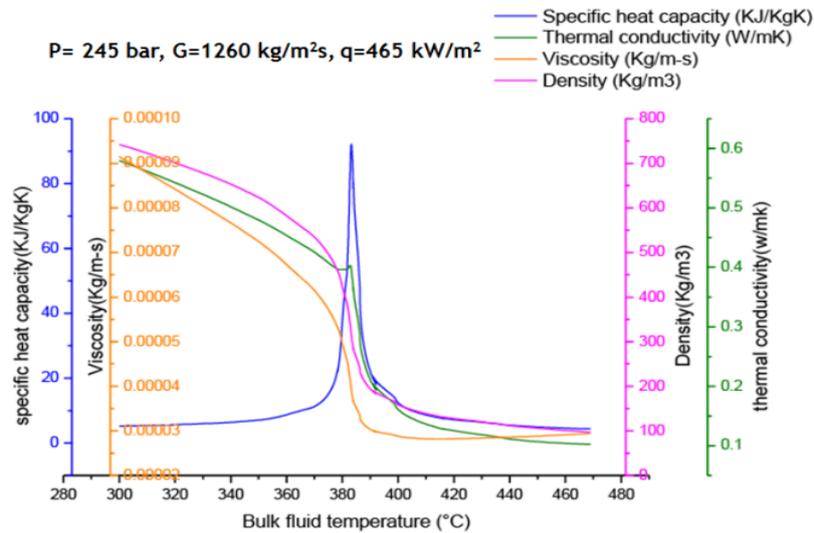


Figure 1. Variation of properties near pseudo critical point

The efficiency of a standard subcritical power plant is in the range 36-38% while that of a supercritical plant is above 42% and for ultra-supercritical and advanced ultra-supercritical plants, the value is above 46%. The increase in pressure to improve efficiency was the only scope of improvement left out in the Rankine cycle as for other limits, no further pushing was possible nor they were economical. The ever increasing quest for efficiency improvement has seen more power plants being designed and operated at higher pressures. The initial sluggishness in the commercial scale development of the supercritical plants was overcome by the development of higher grade austenitic steels. By late 80's, several supercritical power plants were engineered and started commercial operation with superheated steam conditions as high as 600°C and 300 bar. Research in this area started picking up in Asia by the early 21st century. China was the front-runner among the Asian countries followed by India. Because of the advantages like higher unit efficiency, lower pollutant emission and lower fuel consumption, supercritical boilers outperform the subcritical power plants. The difficulty in meeting the fuel supply to demand and the awareness of global climate shift has made this cleaner technology more acceptable among developing countries like India.

It is found that the temperature of the fluid (water) in the evaporator region of the present supercritical boilers is roughly in the range of 350° to 450°C, within which the pseudo critical temperature of 384.7 °C at 250 bar is likely to occur. In the vicinity of the pseudo critical point, strong variations in properties combined with a high heat flux can lead to a Deteriorated Heat Transfer (DHT). Even though this phenomenon is not as severe as the similar phenomenon in subcritical range namely Departure from Nucleate Boiling (DNB), it can cause appreciable decrease in the heat transfer coefficient and hence increase in the wall temperature. Under extreme conditions, the temperature can exceed the metal oxidation limit and the tube can fail. Therefore, it is indeed necessary to study the heat transfer behavior of water at supercritical conditions and the influence of the same on wall temperature. Since the turbulence modelling is highly influenced by the large variation in fluid properties at supercritical conditions, mathematical modelling is an uphill task. The turbulence model suitable for a certain range of operation may not give similar results under other conditions.

That is, a model which predicts well the heat transfer deterioration may not predict the heat transfer deterioration well. So it is important to understand various models and how they behave under various conditions. A set of standard two equation models (Standard $k-\epsilon$, RNG $k-\epsilon$, $k-\omega$ SST) and 5 Low Reynolds Number (Low Re) models are chosen in the present study for evaluation using experimental results of Yamagata *et al.* (1972).

II. NUMERICAL METHOD

2.1 Turbulence models

Turbulence flows are high Reynolds number flows which are chaotic or random in nature. The flow field parameters of a turbulent flow are highly unpredictable and a fully deterministic or analytical solution is difficult since the parameters cannot be expressed quantitatively with accuracy. So turbulent flows are described statistically. The unsteady nature of the parameters (velocity, temperature and pressure) assists in the transfer of matter, momentum and energy between fluid particles. The diffusivity of turbulence causes rapid mixing and increased rates of momentum, heat, and mass transfer. A flow that looks random but does not exhibit the spreading of velocity fluctuations through the surrounding fluid is not turbulent. The velocity and pressure components are decomposed into the sum of a mean component and a fluctuating component. The mean component can be time average or space average, but mostly time average. The resultant equations are called the Reynolds Averaged Navier Stokes (RANS) equations.

Expressing as the sum of mean and fluctuating component,

$$n_i = \bar{n}_i + \acute{n}_i \quad (1)$$

where n_i is the specific property which can be velocity, temperature or density,

\bar{n}_i is the time average component and \acute{n}_i is the turbulence fluctuating component

Thus the process of averaging has introduced a new term (product of fluctuating components) known as the Reynolds stress or turbulent stress. As a result of this there are six additional unknowns in the RANS equations. The main task of turbulence modelling is to develop computational procedures of sufficient accuracy to accurately predict the Reynolds stresses and the scalar transport terms. This will then allow for the computation of the time averaged flow and scalar fields without having to calculate the actual flow fields over long time periods. This process is called closure of the model. The additional turbulent stress term in the RANS is predicted with required accuracy in a number ways. Two equation models are the most widely used models due to their simplicity. These models provide expression for computation of kinetic energy k and for the turbulence length scale or equivalent. Apart from the standard two equation models, there are Low Reynolds number models which consider viscous damping functions in the viscous sub layer where the Reynolds number is quite low. They are modified $k-\epsilon$ models.

2.2 Governing Equations

As in any numerical code for flow dynamics and heat transfer, the governing equations are simply versions of the conservation laws of classic physics-

2.2.1 Conservation of Mass (Continuity Equation): For incompressible flow, continuity equation for the mean component can be written as

$$\frac{\partial \bar{u}_j}{\partial x_j} = 0 \quad (2)$$

2.2.2 Conservation of Momentum: The general form of momentum equation can be written as

$$\rho \left[\frac{\partial \bar{u}_i}{\partial t} + \frac{\partial (\bar{u}_i \bar{u}_j)}{\partial x_j} \right] = - \frac{\partial \bar{p}}{\partial x_i} + \frac{\partial}{\partial x_j} \left(\mu \frac{\partial \bar{u}_i}{\partial x_j} \right) - \frac{\partial (\rho \bar{u}_i \acute{u}_j)}{\partial x_j} \quad (3)$$

2.2.3 Conservation of Energy: The general form of energy equation can be written as

$$\rho \frac{\partial k}{\partial t} + \rho \bar{u}_j \frac{\partial k}{\partial x_j} = -\tau_{ij} \frac{\partial \bar{u}_i}{\partial x_j} - \rho \epsilon + \frac{\partial}{\partial x_j} \left[\left(\mu + \frac{\mu_\tau}{\sigma_k} \right) \frac{\partial k}{\partial x_j} \right] \quad (4)$$

where τ_{ij} is the turbulence stress tensor, ϵ is the dissipation rate of k .

2.3 Physical Model

In the present work, a two-dimensional axisymmetric model is developed to numerically study the heat transfer characteristic of supercritical water flowing through a vertical tube with uniform heating and the effect of diameter on heat transfer has been analysed. A 7.5 mm diameter tube is used for analysis. Length of the tube

is varied from case to case to obtain the required outlet temperature. In order to obtain a fully developed flow, the modelled tube is extended at the inlet by 0.5m. The physical model is depicted in Figure 2.

For the calculation of the thermodynamic properties of supercritical fluids, the NIST database available in FLUENT is used. The SIMPLE algorithm is used to simultaneously solve the velocity and pressure equations. The QUICK scheme is used to approximate the flow field variables in the discretized convective terms in transport equations.

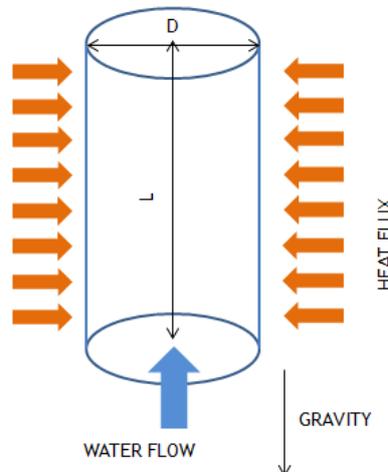


Figure 2. Schematic representation of the tube geometry

2.4 Grid Independence Test

The selection of grid is important since the accuracy of the results depends on the fineness of the grid. On the other side, beyond a certain point, further refinement of grid will not improve the solution; rather it will turn out to be wastage of computational time. Here, a grid independence test has been conducted to choose an appropriate mesh for the above physical model. Grids with 100 x 600, 120 x 600, 140 x 600, 175 x 600 and 200 x 600 (radial nodes x axial nodes) were tested. Non-uniform nodes with a successive ratio of 1.025 in the radial direction and uniform nodes in the axial direction have been used for all the grids. In order to evaluate the accuracy of the test results, the experimental data of Yamagata et al. (1972) has been selected for validation. A 7.5 mm diameter tube is used with a heat flux of 465 kW/m² and mass flux of 1260 kg/m². Operating pressure was 245 bar. The comparison is plotted in Figure 3. It can be seen that 200 x 600 grids matches well with that of experiment data. Further increase in grids does not improve the accuracy of the solution.

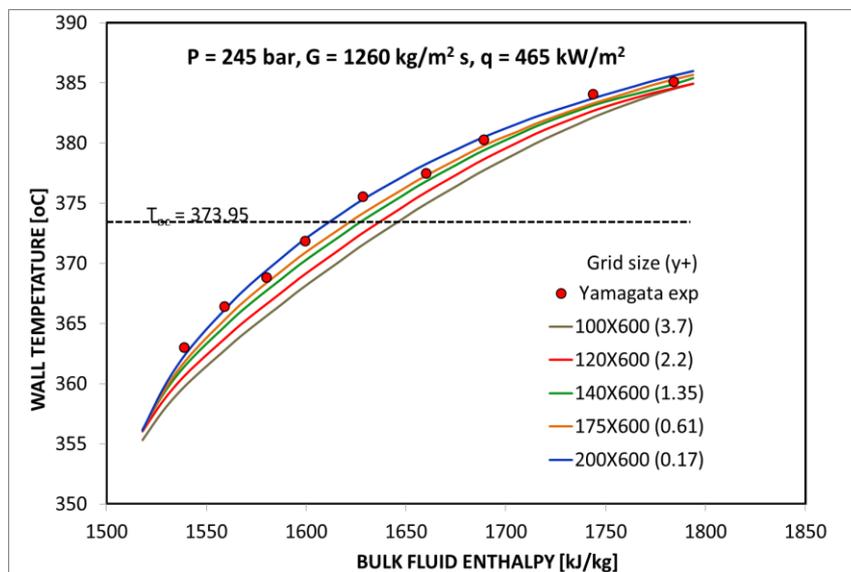


Figure 3. Grid Independence test

III. RESULTS AND DISCUSSION

3.1 Model comparison

Eight two equation models available in the FLUENT are chosen for heat transfer prediction capability study in the present case. Standard two equation turbulence models such as standard $k-\epsilon$, RNG $k-\epsilon$ model with enhanced wall treatment and $k-\omega$ SST are chosen. Five low Reynolds number models namely models by Abid, Lam-Bremhost, Launder-Sharma, Yang-Shih and Chang-Hsieh-Chen are also selected. A 7.5 m tube is used with an inlet mass flux of 1260 kg/m^2 and heat flux of 233 kW/m^2 has been used. The heat flux to mass flux ratio is less in this case and normal heat transfer is expected throughout the operation range with an improvement in heat transfer near the pseudo critical region. For comparing the heat transfer prediction capability of the models, the plot of tube wall temperature vs. the bulk fluid enthalpy is chosen. This is a good tool for analysis since it can be inferred from the plot that the region under consideration favours heat transfer enhancement or deterioration. A decrease in the slope of the wall temperature curve depicts an improvement in heat transfer and an increase in slope indicates the other way around. The wall temperature plot is compared with the experimental result of Yamagata et al. (1972) under same conditions for all eight models selected for evaluation.

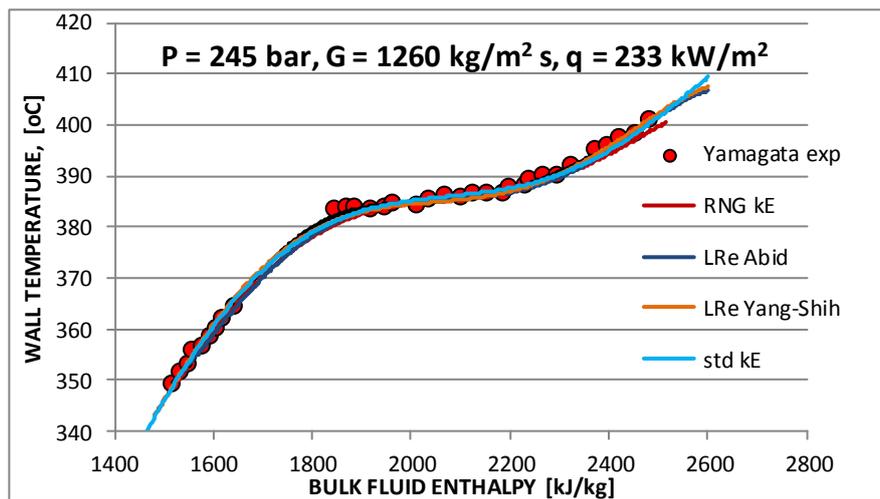


Figure 4. Models predicting the wall temperature closely

It is seen that four out of eight models considered are closely following the experimental results as seen in Figure 4. The average difference between the wall temperature between these models and the experimental data is less than 1°C . For better clarity, a portion of the above plot is expanded to have a closer look (Figure 5). A region close to pseudo critical region is selected since the property variations are maximum in this region. Due to the property variation, the deviation from the experimental data is expected to be a maximum in this region. So the capability of the models to predict the heat transfer characteristics in the above region is particularly checked.

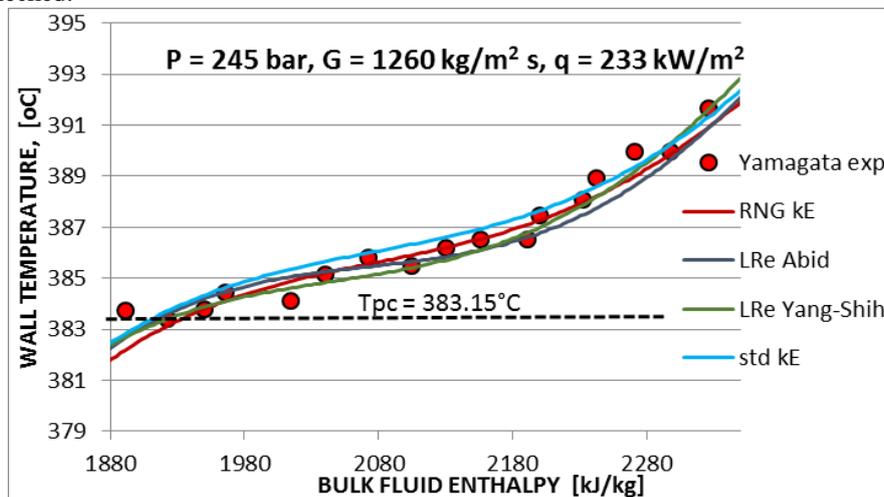


Figure 5. Expanded view of models predicting the wall temperature closely

It is seen that the RNG $k-\epsilon$ model and the Low Reynolds number model by Abid are the closest to experimental results. The average differences calculated between the wall temperature prediction by these models and the experimental wall temperature are 0.3°C and 0.4°C respectively. It can be seen that the wall temperatures of all the models follow similar trend. When the bulk enthalpy is less than 1900 kJ/kg, the wall temperature ascends linearly following enthalpy increase. The wall temperature curves become flat between 1900 kJ/kg and 2200 kJ/kg. This is because the region covers the pseudo critical zone where the specific heat is substantially higher than the normal value. The specific heat increases from the normal value of 4.18 kJ/kg-K and assumes a value as high as few tens (90 kJ/kg-K) as seen in Figure 1. This enhances the heat transfer coefficient and the fluid enthalpy increases without notable increase in the wall temperature. The wall temperature increases again beyond 2200 kJ/kg where specific heat falls down. So it can be said that the above 4 models are working well in the supercritical region under the heat transfer enhancement region.

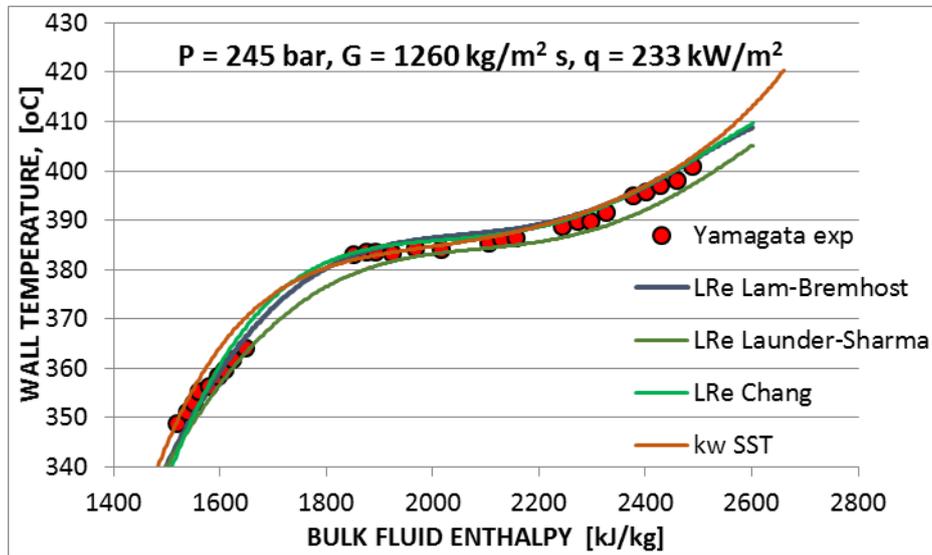


Figure 6. Wall temperature prediction by models

The wall temperature plots for the other four models are slightly offset from the experimental data with the average difference between the prediction and experimental wall temperature ranging from 1.1 °C to 2.7 °C. The wall temperature plots are depicted in Figure 6. The average difference between the wall temperature prediction by the models and the experimental wall temperature is tabulated in Table 1 below

Sl no	Model	Average difference in temperature [°C]
1	RNG $k-\epsilon$ model with enhanced wall treatment	0.3
2	Low Re model by Abid	0.4
3	Low Re model by Yang-Shih	0.5
4	Standard $k-\epsilon$ model	0.5
5	$k-\omega$ SST model	1.1
6	Low Re model by Lam-Bremhost	1.7
7	Low Re model by Chang-Hseih-Chen	1.9
8	Low Re model by Launder-Sharma	2.7

It is seen that first four models (RNG $k-\epsilon$ model with enhanced wall treatment, Low Re model by Abid, Low Re model by Yang-Shih, Standard $k-\epsilon$ model) are predicting closely the wall temperature compared to the experimental data. The average temperature difference is within 1 °C. The latter four models ($k-\omega$ SST model, Low Re model by Lam-Bremhost, Low Re model by Chang-Hseih-Chen, Low Re model by Launder-Sharma) are showing higher deviations up to 2.7°C.

Computation time is also considered in the model selection. Low Re models are more time consuming compared to the standard models. They take approximately 20% more time than the standard model. So considering the accuracy of prediction and the computational time, RNG $k-\epsilon$ model with enhanced wall treatment is the best model under the conditions considered.

3.2 Validation

The heat flux considered in the above tests was 233 kW/m^2 . The above study was extended to a higher heat flux (698 kW/m^2) to ascertain the capability of selected RNG $k-\epsilon$ model at higher ranges. The wall temperature prediction is compared with the experimental data of Yamagata et. al (1972) under the same conditions. The plot is shown in Figure 7. It is seen that at higher heat flux also the model is predicting well. The average temperature difference between prediction and experiment data is within 0.5°C .

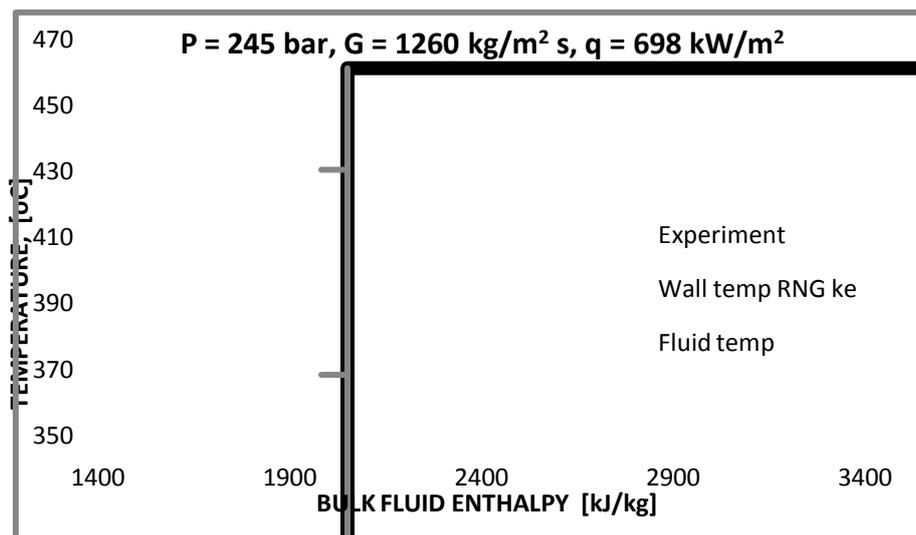


Figure 7. Wall temperature prediction by RNG $k-\epsilon$ model with EWT

IV. CONCLUSION

- For the selected case, the standard two equation models such as standard $k-\epsilon$, RNG $k-\epsilon$ model with enhanced wall treatment and $k-\omega$ SST models are in good agreement with the experimental results under supercritical conditions with low heat flux to mass flux ratios.
- Low Re models such as models by Abid, Lam-Bremhost, Launder-Sharma, Yang-Shih and Chang-Hsieh-Chen also work well under the same conditions.
- Under low heat flux to mass flux ratios the maximum difference between the predicted wall temperature and experimental data is 2.7°C for the model by Launder-Sharma which is the maximum deviation.
- RNG $k-\epsilon$ model with enhanced wall treatment is the best among the compared models with respect to the accuracy and the computational time. The average temperature difference for this model is 0.3°C .
- RNG $k-\epsilon$ model with enhanced wall treatment works well at moderately high heat flux of 698 kW/m^2 and for an enthalpy range 1450 kJ/kg to 2750 kJ/kg .

REFERENCES

- [1] K Yamagata, K Nishikawa, S Hasegawa, I Fuji and S Yoshida. Forced Convective Heat Transfer to Supercritical Water Flowing in Tubes, Int. J. Heat Mass Transfer, 15, pp. 2575-2593. 1972
- [2] Medhat Sharabi and Walter Ambrosini. Discussion of heat transfer phenomena in fluids at supercritical pressure with the aid of CFD models, Annals of Nuclear Energy, 36, pp. 60-71, 2009
- [3] Pioro I.L., Duffey R.B. Experimental heat transfer in supercritical water flowing inside channels (survey). Nucl. Eng. Des. 235, 2407-2430, 2005
- [4] Incropera F.P., DeWitt D.P., Bergman Th.L., Lavine A.S. Fundamentals of Heat and Mass Transfer, 6th edition. J. Wiley & Sons, New York, NY, USA, 2007
- [5] NIST Reference Fluid Thermodynamic and Transport Properties. REFPROP, NIST Standard Reference Database 23 (on diskette: Executable with Source), Ver. 7.0, E.W. Lemmon, M.O. McLinden, M.L. Huber, U.S. Dept. of Commerce, 2002
- [6] Xiaojing Zhu, Qincheng Bi, Dong Yang, Tingkuan Chen. An investigation on heat transfer characteristics of different pressure steam-water in vertical upward tube, Nuclear Engineering and Design 239, pp. 381-388, 2009

Moments of the Reliability, $R = P(Y < X)$, As a Random Variable

Mohammed A. Shayib and Aliakbar Montazer Haghighi

Department of Mathematics
Prairie View A&M University
Prairie View, Texas 77446 USA

ABSTRACT

The estimation of the reliability $R = P(Y < X)$, when X and Y are independent and identically distributed random variables, has been extensively studied in the literature. Distribution functions such as Normal, Burr Type X, Exponential, Gamma, Weibull, Logistic, and Extreme Value have been considered. The expressions to estimate R have been expressed as function of positive parameters from the assumed distributions. No article yet has considered the expression for R as a random variable by itself. In this paper we will do so and consider some distributions that have been introduced in the literature to estimate it. The objective of this paper is to derive the probability density function (pdf) of R , under the assumption that the parameter of the distribution under investigation is itself a random variable with some specified pdf. Once the pdf of R is derived, its moments and other properties are investigated. Several distributions including the Exponential, Chi-Square, Gamma, Burr Type X and Weibull are assumed for this research and are investigated. Furthermore, two cases are considered for each distribution. It has been found that the moment generating function techniques for finding the moments do not work.

I. INTRODUCTION AND BACKGROUND

The product reliability seems, finally, becoming the top priority for the third millennium and a technically sophisticated customer. Manufacturers and all other producing entities are sharpening their tools to satisfy such customer. Estimation of the reliability has become a concern for many quality control professionals and statisticians. Let Y represent the random value of a stress that a device (or a component) will be subjected to in service and X represent the strength that varies from product to product in the population of devices. The device fails at the instant that the stress applied to it exceeds the strength and functions successfully whenever Y is less than X . Then, the reliability (or the measure of reliability) R is defined as $P(Y < X)$, that is, the probability that a randomly selected device functions successfully. Algebraic forms of R , for different distributions, have been studied in the literature. Among those distributions considered are the normal, exponential, gamma, Weibull, Pareto, and recently extreme value family by Nadarajah (2003). Haghighi and Shayib (2009) have considered the logistic and extreme value distributions. In all cases, X and Y are assumed to be independent random variables. If $H_Y(y)$ and $f_X(x)$ are the cumulative distribution function (cdf) and the probability density function (pdf) of Y and X , respectively, then it is well known that

$$R = P(Y < X) = \int_{-\infty}^{\infty} H_Y(z) f_X(z) dz. \quad (1)$$

The estimation of R is a very common concern in statistical literature in both distribution-free and parametric cases. Enis and Geisser (1971) studied Bayesian approach to estimate R . Different distributions have been assumed for the random variables X and Y . Downton (1973) and Church and Harris (1970) have discussed the estimation of R in the normal and gamma cases, respectively. Studies of stress-strength model and its generalizations have been gathered in Kotz et al. (2003). Gupta and Lvin (2005) studied the monotonicity of the failure rate and mean residual life function of the generalized log-normal distribution.

Out of 12 different forms of cdf Burr (1942) introduced for modeling survival data, Burr Type X and Burr Type XII distributions have received an extensive attention. Awad and Gharraf (1986) used the Burr Type X model to simulate a comparison of three estimates for R , namely, the minimum variance unbiased, the maximum likelihood, and the Bayes estimators. They also studied the sensitivity of the Bayes estimator to the prior parameters. Ahmad et al. (1997) used the same assumptions except they further assumed that the scale

parameters are known. Then, they used maximum likelihood (MLE), Bayes and empirical techniques to deal with the estimation of R in such a case. They compared the three methods of estimation using the Monte-Carlo simulation. Additionally, they presented a comparison among the three estimators and some characterization of the distribution. The first characteristic they studied was based on the recurrence relationship between two successively conditional moments of a certain function of the random variable, whereas the second was given by the conditional variance of the same function. Surles and Padgett (1998) introduced the scaled Burr Type X distribution and name it as Scaled Burr Type X or the Generalized Rayleigh Distribution (GRD). They considered the inference on R when X and Y are independently distributed Burr Type X random variables, discussed the existing and new results on the estimation of R , using MLE, and introduced a significance test for R . They also presented Bayesian inference on R when the parameters are assumed to have independent gamma distribution. They, further, offered an algorithm for finding the highest posterior density interval for R and using different methods of estimation, calculated the value of R under the assumed distributions.

Baklizi and Abu-Dayyeh (2006) have considered the problem of estimating $R = (Y < X)$ when X and Y have independent exponential distributions with parameters θ and λ , respectively. Also, Baklizi and El-Masri (2004) had considered the same problem as in Baklizi and Abu Dayyeh (2006), but with a common location parameter μ . Surles and Padgett (1998) considered the inference on R when X and Y are independent and identically distributed as Burr-type- X random variables with parameters θ and λ , respectively. In addition, Raqab and Kundu (2009) considered the scaled Burr Type X distribution for comparing different estimators of R . Moreover, the expression for R came out as is specified in

$$R = P(Y < X) = \frac{1}{1 + \frac{\lambda}{\theta}}, \tag{2}$$

where the parameter λ is of the stress Y and θ is that of the strength X . It is to noted here that R increases as the ratio λ/θ decreases, and, thus, the parameter λ of the stress Y needs to be much smaller than θ of the strength X . Shayib (2005) had considered the effect of the sample size and the ratio between the parameters on the estimation of R , when X and Y shared the Burr Type X distribution. Haghghi sand Shayib (2008), considered the Weibull distribution when estimating R . A common application of the Weibull distribution is to model the lifetimes of components such as bearings, ceramics, capacitors, and dielectrics. Haghghi sand Shayib (2009) also considered the estimation of R in the case of the logistic distribution.

Generally, the *cdf* and *pdf* of a two-parameter logistic distribution are given, respectively, as:

$$F_X(x) = \frac{1}{1 + e^{-\frac{x - \alpha}{\beta}}} \text{ and } f_X(x) = \frac{e^{-\frac{x - \alpha}{\beta}}}{\beta \left(1 + e^{-\frac{x - \alpha}{\beta}}\right)^2}, \quad -\infty < x, \alpha < \infty, \beta > 0, \tag{3}$$

where α and β are the location and scale parameters, respectively. We consider the unknown-one-parameter logistic distributions for both X and Y , that is, we assume that the location parameter α is 0. Hence, *cdf* and *pdf* are, respectively, as follow:

$$H_Y(y) = \frac{1}{1 + e^{-\frac{y}{\beta_1}}} \text{ and } h_Y(y) = \frac{e^{-\frac{y}{\beta_1}}}{\beta_1 \left(1 + e^{-\frac{y}{\beta_1}}\right)^2}, \quad -\infty < y < \infty, \beta_1 > 0, \tag{4}$$

$$F_X(x) = \frac{1}{1 + e^{-\frac{x}{\beta_2}}} \text{ and } f_X(x) = \frac{e^{-\frac{x}{\beta_2}}}{\beta_2 \left(1 + e^{-\frac{x}{\beta_2}}\right)^2}, \quad -\infty < x < \infty, \beta_2 > 0. \tag{5}$$

Haghghi sand Shayib (2009) proved the following Theorem:

Theorem: When X and Y are independent random variables with logistic probability distribution and the location parameter $\alpha = 0$, then

$$R(\beta_1, \beta_2) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{(-1)^{n+m}}{1 + \frac{m}{n+1} \frac{\beta_2}{\beta_1}}. \tag{6}$$

They also considered the estimation of R in the case of the extreme value distribution. They assumed that both X and Y are independent and each has an unknown-one-parameter extreme-value type 1 (or double exponential or Gumbel-type) distribution. Without loss of generality, it can be assumed that the location parameter is zero for each of X and Y . Hence, in this case, the *cdf* and *pdf* for Y and X are, respectively, as follows:

$$F_Y(y; \theta_1) = e^{-\frac{y}{\beta_1}} \text{ and } f(y; \theta_1) = \frac{1}{\theta_1} e^{-\frac{y}{\beta_1}} e^{-e^{-\frac{y}{\beta_1}}}, \quad -\infty < y < \infty, \theta_1 > 0, \tag{7}$$

and

$$H_X(x; \theta_2) = e^{-e^{-\frac{x}{\theta_2}}} \text{ and } h(x; \theta_2) = \frac{1}{\theta_2} e^{-\frac{x}{\theta_2}} e^{-e^{-\frac{x}{\theta_2}}}, \quad -\infty < x < \infty, \theta_2 > 0. \tag{8}$$

Moreover, Haghghi sand Shayib (2009) proved the following Theorem:

Theorem: Given (7) and (8), if

$$k \frac{\theta_2}{\theta_1} < 1, \quad k = 0, 1, 2, \dots,$$

then,

$$R(\theta_1, \theta_2) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \Gamma\left(k \frac{\theta_2}{\theta_1} + 1\right). \tag{9}$$

II. R AS A RANDOM VARIABLE

We have seen that R has been expressed as a ratio of one positive parameter over the sum of two positive parameters. This ratio has appeared under the assumption of exponential, Burr Type X, gamma, and Weibull distributions. In the cases of logistic and extreme-value distributions, an infinite series presentation of R has been derived; see Haghghi sand Shayib (2009). Thus, in the remaining of the sequel we will assume

$$R = \frac{U}{U + V}, \tag{10}$$

where $U > 0, V > 0$ and that U and V are independent random variables.

We will consider the following distribution functions: the exponential, Chi-Square, gamma, Burr Type X, and Weibull. Without loss of generality and for simplifying the presentation, the one parameter in the considered distributions will be denoted by β . As a special case of the gamma distribution, we will consider the distributions that correspond to the exponential and Chi Square. This will be dealt with in section 3 for the exponential distribution as the underlying assumption for the variables that are involved in R . The Chi-Square distribution will be taken as the underlying one in section 4. In Section 5, we will investigate a special case of the gamma distribution. Conclusion remarks are summarized in the last section.

III. EXPONENTIAL DISTRIBUTIONS FOR U AND V

We will assume that U and V are independent and identically distributed random variables that share an exponential distribution, but with different parameters, namely,

$$U \sim Exp(\beta_1) \text{ and } V \sim Exp(\beta_2),$$

where β_1 and β_2 are positive. In this case the probability density functions are known as:

$$f_1(u) = \begin{cases} \frac{1}{\beta_1} e^{-\frac{u}{\beta_1}}, & u > 0 \\ 0, & u \leq 0, \end{cases} \quad (11)$$

and

$$f_2(v) = \begin{cases} \frac{1}{\beta_2} e^{-\frac{v}{\beta_2}}, & v > 0 \\ 0, & v \leq 0, \end{cases} \quad (12)$$

Since U and V are independent, their joint probability density function will be

$$g(u, v) = \begin{cases} \frac{1}{\beta_1 \beta_2} e^{-\left(\frac{u}{\beta_1} + \frac{v}{\beta_2}\right)} & u, v > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

In this case, $g(u, v) > 0, \forall(u, v)$ such that $u > 0$ and $v > 0$. We are interested in R as a function of U and V in the form (10).

Let us define

$$R_1 = \frac{u}{u + v} = h_1(u, v) \text{ and } R_2 = u + v = h_2(u, v). \quad (14)$$

This choice of R_2 yields an inverse transformation of the following form

$$U = R_1 R_2 = h_1^{-1}(R_1, R_2) \text{ and } V = R_2(1 - R_1) = h_2^{-1}(R_1, R_2). \quad (15)$$

The Jacobian of this transformation will be

$$J = \det \begin{pmatrix} R_2 & R_1 \\ -R_2 & 1 - R_1 \end{pmatrix} = R_2. \quad (16)$$

Thus, the joint probability density function of R_1 and R_2 is given by

$$f(R_1, R_2) = \begin{cases} \frac{R_2}{\beta_1 \beta_2} e^{-\left[\frac{\beta_2 R_1 R_2 + \beta_1 R_2(1 - R_1)}{\beta_1 \beta_2}\right]}, & R_1, R_2 > 0, R_2(1 - R_1) > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

From (17), the marginal density functions for R_1 and R_2 . For instance, for R_1 we have

$$h(R_1) = \frac{\beta_1 \beta_2}{[(\beta_2 - \beta_1)R_1 + \beta_1]^2}, \quad 0 < R_1 < 1, \beta_1, \beta_2 > 0. \quad (18)$$

Note that when the two parameters are equal $h(R_1) = 1$, and R_1 is uniformly distributed over $(0, 1)$. This situation can arise when the stress and strength, that the component is exposed to, share the same density function that follow the exponential, Burr Type X, or the Weibull distribution. This in turn shows the result that was calculated by Shayib (2005) that the value of R is $1/2$, and the first moment will not exceed that. Moreover, when the two parameters are equal, then R_2 will have a gamma density with parameters $\alpha = 2$ and β , the common parameter of the assumed distributions.

Now after establishing the probability density function for R , we will find its moments. Relation (18) is the density function of R . Thus, we have

$$\begin{aligned}
 E(R) = \mu &= \int_0^1 r h(r) dr = \beta_1 \beta_2 \int_0^1 \frac{r}{[(\beta_2 - \beta_1)r + \beta_1]^2} dr \\
 &= \frac{\beta_1 \beta_2}{(\beta_2 - \beta_1)^2} \left[\ln \left(\frac{\beta_2}{\beta_1} e^{-\frac{\beta_2 - \beta_1}{\beta_2}} \right) \right], \quad \frac{\beta_2}{\beta_1} e^{-\frac{\beta_2 - \beta_1}{\beta_2}} > 1. \quad (19)
 \end{aligned}$$

From (19) we will have

$$[E(R)]^2 = \mu^2 = \frac{\beta_1^2 \beta_2^2}{(\beta_2 - \beta_1)^4} \left[\ln \left(\frac{\beta_2}{\beta_1} e^{-\frac{\beta_2 - \beta_1}{\beta_2}} \right) \right]^2 \quad (20)$$

and from (18) we have

$$E(R^2) = \int_0^1 r^2 h(r) dr = \beta_1 \beta_2 \int_0^1 \frac{r^2}{[(\beta_2 - \beta_1)r + \beta_1]^2} dr. \quad (21)$$

Thus, the variance of R can be found from (20) and (21).

Subsequently, other moments can be found. It is to be noted here that the Moment Generating Function technique to find the moments is a forbidden task.

IV. CHI-SQUARE DISTRIBUTION FOR BOTH U AND V

In this section, we will assume that U and V are iid random variables that share the Chi square distribution when the parameters are different, i.e., with degrees of freedoms d_1 and d_2 , $d_1 \neq d_2$. It can be shown that the pdf of R will have a Beta density function with parameters $d_1/2, d_2/2$, that is

$$g(r) = \begin{cases} r^{\left[\frac{d_1}{2}-1\right]} (1-r)^{\left[\frac{d_2}{2}-1\right]} \frac{\Gamma\left(\frac{d_1}{2} + \frac{d_2}{2}\right)}{\Gamma\left(\frac{d_1}{2}\right) \Gamma\left(\frac{d_2}{2}\right)}, & 0 < r < 1, \\ 0, & \text{otherwise.} \end{cases} \quad (22)$$

From (22), the mean and the variance are given by

$$\mu = \frac{d_1}{d_1 + d_2} \quad (23)$$

and

$$\sigma^2 = \frac{2d_1 d_2}{(d_1 + d_2 + 2)(d_1 + d_2)^2}, \quad (24)$$

respectively.

When the two Chi square distributions share the same degrees of freedom, then R will have a Beta distribution with parameters $d/2$ and $d/2$, with the following probability density function:

$$g(r) = \begin{cases} r^{\left[\frac{d}{2}-1\right]} (1-r)^{\left[\frac{d}{2}-1\right]} \frac{\Gamma\left(\frac{d}{2} + \frac{d}{2}\right)}{\Gamma\left(\frac{d}{2}\right) \Gamma\left(\frac{d}{2}\right)}, & 0 < r < 1, \\ 0, & \text{otherwise.} \end{cases} \quad (25)$$

In this case, the mean and variance of R are well known and are given by

$$\mu = \frac{1}{2} \text{ and } \sigma^2 = \frac{1}{4d + 4},$$

see Hogg and Tanis (1988), for instance.

We note that the moment generating function for the Beta distribution cannot be expressed in a close form. Thus, to find other moments, we will appeal to integration methods.

V. GAMMA DISTRIBUTION FOR BOTH U AND V

Although the gamma distribution has two identifying parameters, by fixing one at a time, two more distributions will emerge. Using the notation that if the random variable X has a gamma distribution with parameters α and β , that is, $X \sim G(\alpha, \beta)$. It is well known that when $\alpha = 1$, G will be the exponential distribution with parameter β .

Now, in G we assume $\beta = 1$. Thus, if U and V , each has a gamma distribution with the same parameter, say α , we see that the pdf for both U and V will, respectively, be

$$f_1(u) = \begin{cases} \frac{1}{\Gamma(\alpha)} u^{\alpha-1} e^{-u}, & 0 < u < \infty, \\ 0, & \text{otherwise.} \end{cases} \tag{26}$$

and

$$f_2(v) = \begin{cases} \frac{1}{\Gamma(\alpha)} v^{\alpha-1} e^{-v}, & 0 < v < \infty, \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

Since U and V are independent, it is clear that their joint density function will be

$$h(u, v) = \begin{cases} \frac{1}{[\Gamma(\alpha)]^2} (uv)^{\alpha-1} e^{-(u+v)}, & 0 < u, v < \infty, \\ 0, & \text{otherwise.} \end{cases} \tag{28}$$

Thus, in this case the joint density functions of R_1 and R_2 is given by

$$h(r_1, r_2) = \begin{cases} \frac{r_1^{\alpha-1} r_2^{2\alpha-1} (1 - r_1)^{\alpha-1}}{[\Gamma(\alpha)]^2} e^{-r_2}, & r_2 > 0, 0 < r_1 < 1, \\ 0, & \text{otherwise.} \end{cases}$$

In accordance with Theorem 1, Section 2.4 of Hogg and Craig (1995), the random variables R_1 and R_2 are independent. The Marginal pdf of R_1 is given by

$$g(r_1) = \begin{cases} \frac{r_1^{\alpha-1} (1 - r_1)^{\alpha-1} \Gamma(2\alpha)}{[\Gamma(\alpha)]^2}, & 0 < r_1 < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{29}$$

This pdf is a Beta distribution with parameters α and α . Thus, the mean and variance of R are given, respectively, by

$$\mu = \frac{1}{2} \text{ and } \sigma^2 = \frac{1}{4\alpha + 4}, \tag{30}$$

When the parameters of the assumed gamma distributions are different, that is, with α_1 and α_2 , the pdf of R will take the following form:

$$g(r_1) = \begin{cases} r_1^{[\frac{\alpha_1}{2}-1]} (1 - r_1)^{[\frac{\alpha_2}{2}-1]} \frac{\Gamma\left(\frac{\alpha_1}{2} + \frac{\alpha_2}{2}\right)}{\Gamma\left(\frac{\alpha_1}{2}\right) \Gamma\left(\frac{\alpha_2}{2}\right)}, & 0 < r_1 < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{31}$$

That is, R will have a Beta distribution with parameters $\alpha_1/2$ and $\alpha_2/2$. Hence, the mean and the variance are given by

$$\mu = \frac{\alpha_1}{\alpha_1 + \alpha_2} \text{ and } \sigma^2 = \frac{2\alpha_1\alpha_2}{(\alpha_1 + \alpha_2 + 2)(\alpha_1 + \alpha_2)^2}, \tag{32}$$

see Hogg and Tanis 1988.

Note that the moment generating function for the Beta distribution cannot be expressed in a close form for finding other moments.

VI. WEIBULL DISTRIBUTION FOR U AND V

In this section we consider the case that the distributions of U and V are independent Weibull random variables, sharing the same known shape parameter α , but unknown and different scale parameters, namely, $U \sim \text{Weibull}(\alpha, \beta_1)$ and $V \sim \text{Weibull}(\alpha, \beta_2)$. The Weibull distribution is commonly used for a model for life lengths and in the study of breaking strengths of materials because of the properties of its failure rate function. Resistors used in the construction of an aircraft guidance system have life time lengths that follow a Weibull distribution with $\alpha = 2$ and $\beta = 10$ (with measurements in thousands of hours).

In this paper we consider the Weibull distribution with two parameters, namely, α and β , where α is the shape parameter while $\beta^{1/\alpha}$ is the scale parameter of the distribution. The probability density function of the Weibull distribution that we will use based on this setup is

$$f(u) = \begin{cases} \frac{\alpha}{\beta_1} u^{\alpha-1} e^{-\frac{u^\alpha}{\beta_1}}, & u, \alpha, \beta_1 > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{33}$$

$$f(v) = \begin{cases} \frac{\alpha}{\beta_2} v^{\alpha-1} e^{-\frac{v^\alpha}{\beta_2}}, & v, \alpha, \beta_2 > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{34}$$

$$g(u, v) = \begin{cases} \alpha^2 \beta_1 \beta_2 (uv)^{\alpha-1} e^{-(\beta_1 u + \beta_2 v)}, & \alpha, \beta_1, \beta_2, u, v > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{35}$$

Thus, from (14) – (16) we have:

$$g(r_1, r_2) = \begin{cases} \alpha^2 \beta_1 \beta_2 r_1 r_2^{2\alpha-1} (1 - r_1)^{\alpha-1} e^{-r_2^\alpha [\beta_1 r_1^\alpha + \beta_2 (1-r_1)^{\alpha-1}]}, & \alpha, \beta_1, \beta_2, r_2 > 0, 0 < r_1 < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{36}$$

When $\beta_1 = \beta_2 \equiv \beta$, (36) becomes

$$g(r_1, r_2) = \begin{cases} \alpha^2 \beta^2 r_1 r_2^{2\alpha-1} (1 - r_1)^{\alpha-1} e^{-\beta r_2^\alpha [r_1^\alpha + (1-r_1)^{\alpha-1}]}, & \alpha, \beta, r_2 > 0, 0 < r_1 < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{37}$$

Hence,

$$h(r) = \frac{\alpha r^{\alpha-1} (1 - r)^{\alpha-1}}{[r^\alpha + (1 - r)^\alpha]^2}, \quad 0 < r < 1. \tag{38}$$

Note:

- (i) For $\alpha = 1$, certainly, R is uniformly distributed over the interval $(0, 1)$. Hence, the mean and the variance of R , in this case, are, respectively, given by $1/2$ and $1/12$.
- (ii) When $\alpha = 2$, we obtain

$$h(r) = \begin{cases} \frac{2r(1-r)}{[r^2+(1-r)^2]^2}, & 0 < r < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{39}$$

It can be shown that $h(r)$ given in (39) is a pdf for $0 < r < 1$. Hence, the mean and the variance of R are given as $1/2$ and $(\pi - 3)/4$, respectively.

- (iii) In general, for all values of α , we have the following Theorem:

Theorem: The function in (38), for any α , is a pdf over the range of r .

Proof:

It is clear that the function in (38) is positive for all assumed values of r . To prove that the integral of $h(r)$, over the range of r is 1, let $r = \sin^2 t$, $0 \leq t \leq \pi/2$. Then, $dr = 2 \sin t \cos t dt$. Hence, we have the following:

$$\begin{aligned} \int_0^{\pi/2} h(r) dr &= \int_0^{\pi/2} \frac{2\alpha(\sin t)^{2(\alpha-1)} (\cos t)^{2(\alpha-1)} \sin t \cos t}{(\sin^{2\alpha} t + \cos^{2\alpha} t)^2} dt \\ &= \int_0^{\pi/2} \frac{2\alpha(\sin t)^{2\alpha-1} (\cos t)^{2\alpha-1}}{(\sin^{2\alpha} t + \cos^{2\alpha} t)^2} dt \\ &= \int_0^{\pi/2} \frac{2\alpha(\sin t)^{2\alpha-1} (\cos t)^{2\alpha-1}}{\cos^{4\alpha} t (\tan^{2\alpha} t + 1)^2} dt \\ &= \int_0^{\pi/2} \frac{2\alpha(\sin t)^{2\alpha-1}}{\cos^{2\alpha+1} t (\tan^{2\alpha} t + 1)^2} dt \\ &= \int_0^{\pi/2} \frac{2\alpha(\tan t)^{2\alpha-1}}{(\tan^{2\alpha} t + 1)^2} \sec^2 t dt \\ &= - \left[\frac{1}{1 + \tan^{2\alpha} t} \right]_0^{\pi/2} \\ &= 1. \end{aligned}$$

VII. BURR TYPE X DISTRIBUTION FOR BOTH U AND V

We now assume that U and V are iid random variables that share a Burr Type X distribution with a common parameter β . Thus, the density functions of U and V are, respectively, as:

$$f_1(u) = \begin{cases} 2\beta u e^{-u^2} (1 - e^{-u^2})^{\beta-1}, & u, \beta > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{40}$$

and

$$f_2(v) = \begin{cases} 2\beta v e^{-v^2} (1 - e^{-v^2})^{\beta-1}, & v, \beta > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{41}$$

Hence, the joint density is,

$$g(u, v) = \begin{cases} 4\beta^2 u v e^{-(u^2+v^2)} [(1 - e^{-u^2})(1 - e^{-v^2})]^{\beta-1}, & u, v, \beta > 0, \\ 0, & \text{otherwise,} \end{cases} \quad (42)$$

Once again, note that from (14) and (15), there are many values of (u, v) that will give the same value for R_1 .

From (16) we have the joint probability density functions of R_1 and R_2 as

$$g(r_1, r_2) = \begin{cases} 4\beta^2 r_1 r_2^3 (1 - r_1) e^{-[r_1^2 r_2^2 + r_2^2 (1-r_1)^2]} \left[(1 - e^{-r_1^2 r_2^2}) (1 - e^{-r_2^2 (1-r_1)^2}) \right]^{\beta-1}, & 0 < r_1 < 1, r_2 > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (43)$$

For $\beta = 1$, (43) becomes

$$g(r_1, r_2) = \begin{cases} 4r_1 r_2^3 (1 - r_1) e^{-[r_1^2 r_2^2 + r_2^2 (1-r_1)^2]}, & 0 < r_1 < 1, r_2 > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (44)$$

Thus,

$$h(r) = \begin{cases} \frac{2r(1-r)}{[r^2 + (1-r)^2]^2}, & 0 < r < 1, r_2 > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (45)$$

It is easy to show that (45) is a pdf when $0 < r < 1$. Hence, again $E(R) = 1/2$ and the variance of R is $(\pi - 3)/4$.

When $\beta = 2$, the function in (44), will take the form

$$g(r_1, r_2) = \begin{cases} 16r_1 r_2^3 (1 - r_1) e^{-r_2^2 [r_1^2 + (1-r_1)^2]} \left[1 - e^{-r_1^2 r_2^2} - e^{-r_2^2 (1-r_1)^2} + e^{-r_2^2 [r_1^2 + (1-r_1)^2]} \right], & 0 < r_1 < 1, r_2 > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (46)$$

Letting $K = r_1^2 + (1 - r_1)^2$, then $g(r_1, r_2)$ can be expressed as

$$g(r_1, r_2) = \begin{cases} 16r_1 r_2^3 (1 - r_1) e^{-K r_2^2} - 16r_1 r_2^3 (1 - r_1) e^{-r_2^2 (K+r_1^2)} \\ -16r_1 r_2^3 (1 - r_1) e^{-r_2^2 [K+(1-r_1)^2]} + 16r_1 r_2^3 (1 - r_1) e^{-2K r_2^2}, & 0 < r_1 < 1, r_2 > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (47)$$

Thus, in this case,

$$h(r) = 8r(1-r) \left[\frac{1}{K_1^2} + \frac{1}{(K_1 + r^2)^2} + \frac{1}{[K_1 + (1-r)^2]^2} + \frac{1}{4K_1^2} \right], \quad 0 < r < 1, \quad (48)$$

where $K_1 = r^2 + (1 - r)^2$.

Theorem: The function $h(r)$ in (48) is a pdf over the range of R .

Proof:

It is clear that for $0 < r < 1$, $h(r)$ given in (48) is positive. To show that the integral is 1 we proceed as follows. Let us write $h(r)$ as the following sum:

$$h(r) = g_1(r) + g_2(r) + g_3(r) + g_4(r),$$

where

$$g_1(r) = \frac{8r(1-r)}{K_1^2}, \quad g_2(r) = \frac{8r(1-r)}{(K_1 + r_2^2)^2},$$

$$g_3(r) = \frac{8r(1-r)}{[K_1 + (1-r)^2]^2}, \quad \text{and} \quad g_4(r) = \frac{2r(1-r)}{K_1^2}.$$

It is to be noted that $g_1(r) = 4h(r)$, where $h(r)$ is as expressed in (45). Thus, the integral of $g_1(r)$ over the range of r has 4 as its value. Moreover, $g_4(r) = h(r)$, and, hence, the integral of $g_4(r)$ over the range of r is 1. Now for integrals of $g_2(r)$ and $g_3(r)$, let $x = r - \frac{1}{3}$. Thus, using this change of variable and the partial fractions technique, we will have the integrals of $g_2(r)$ and $g_3(r)$ equal to 2. This shows that the integral of $h(r)$ is 1 and that completes the proof.

VIII. CONCLUSION

Different distributions were considered for the investigation $R = P(Y < X)$, when R was taken as a random variable itself, and the parameters of the assumed distributions were taken as random variables themselves. The ratio of the parameters has a greater effect on the value of R than the sample size involved, especially when the parameters are equal. Even though the stress and the strength share the same distribution, the parameter of the stress has to be very small compared with that of the strength, and, thus, the ratio will contribute positively for the value of R . Two cases were investigated under the taken distributions when the two parameters are equal and when they are not. Regardless of the assumed distribution, the mean of R is $1/2$ when the parameters are equal. The moment generating function technique, to find the moments did not work in the cases that were discussed in this paper. Additional work and investigation to find the moments of R is under way. It should be noted that when the parameters of U and V are equal, integrating their joint pdf is not easy; obviously, it will be more difficult when the parameters are unequal.

BIBLIOGRAPHY

- [1] Ahmad, K. E., Fakhry, M. E. and Jaheen, Z. F. (1997). *Empirical Bayes' estimation of $P(Y < X)$ and characterizations of Burr-Type X model*, Journal of Statistical Planning and Inference, 64, 297-308.
- [2] Awad, A. M. and Gharraf, M. K. (1986). *Estimation of $P(Y < X)$ in the Burr case: a Comparative Study*, Commun. Statist, Simula, 15, 389 – 403.
- [3] Burr, I. W. (1942). *Cumulative frequency distribution*, Ann. Math. Statist, 13, 215 – 232.
- [4] Church, J. D. and Harris, B., 1970, *The estimation of reliability from stress-strength Relationships*, Technometrics, 12, 49-54.
- [5] Downton, F. (1973). *The estimation of $Pr(Y < X)$ in the normal case*, Technometrics, 15, 551-558.
- [6] Gupta and Lvin (2005). *Mathematical and Computer Modeling* (Elsevier), 42, 939 – 946.
- [7] Haghighi, A. M. and Shayib, M. A. (2008). *Shrinkage Estimators for Calculating Reliability, Weibull case*, Journal of Applied Statistical Science, Vol. 17, Issue 2, pp. 219-233.
- [8] Haghighi, A. M. and Shayib, M. A. (2009). *Reliability Computation Using Logistic and extreme -Value Distributions*, International Journal of statistics and economics, (Formerly Known by the "Bulletin of Statistics and economics"), Vol. 4 pp. 54-73.
- [9] Hogg, R. V, and Tanis, E. A. (1988). *Probability and Statistical Inference*, 3rd Edition, MacMillan.
- [10] Kotz, S., Lumelskii, Y., and Pensky, M. (2003). *The Stress-Strength Model and its Generalizations*, World Science Press, Singapore.
- [11] Nadarajah, S. (2003). *Reliability for Extreme Value Distributions*, Mathematical and Computer Modeling, 37, 915 – 922.
- [12] Raqab, Mohammad Z. and Kundu, Debasis (2005). *Comparison of Different Estimators of $P(Y < X)$ for a Scaled Burr Type X Distribution*, Communications in Statistics – Simulation and Computation (Taylor & Francis), 34, pp. 564 – 483
- [13] Shayib, M. A. (2005). *Effects of Parameters and Sample Size on the Estimation of $P(Y < X)$, A Simulation Study, Part I*, JSM Proceedings, 142 – 144.
- [14] Surles, J. G. and Padgett, W.J. (1998). *Inference for $P(Y < X)$ in the Burr Type X Model*, Journal of Applied Statistical Science, 7, 225-238.

The Integrated Locomotive Assignment and Crew Scheduling Problem

Aslı Aksoy¹, Alper Altan¹

¹Uludag University Faculty of Engineering Industrial Engineering Department
Gorukle Campus, 16059, Bursa, Turkey

ABSTRACT:

Crew scheduling is an important part of the railway optimization process, but it is also a difficult problem to solve, due to the size of the problem and the complexity of the real life constraints. In this paper, an overview of the crew scheduling problem is given and crew scheduling problem is integrated to locomotive assignment problem and solved by using mixed integer linear programming model.

KEYWORDS: Crew scheduling, integer programming, locomotive assignment, mathematical modeling.

I. INTRODUCTION

Crew management is a major problem for transportation systems such as airlines, railways and public bus transportation. Advances on scheduling methodologies and decision support systems have been substantial in the last years, but they still need improvement, especially from the computational efficiency and practical point of view. Since the privatization and deregulation of most European railroad companies, it has become increasingly important to reduce the overall costs of operations. Labor cost is one of the largest factors in company expenses, so experienced managers are looking for methods to reduce costs.

Crew management is essential to improve productivity, safety and quality of service in many domains, but especially important for transportation systems. In general, crew management involves the construction of efficient sequences of work and rest periods to meet transportation demands and to satisfy constraints. In particular, crew scheduling has been studied from several points of view such as network flow models, mathematical programming, heuristics and dynamic programming.

There are three main approaches for the crew scheduling problem [1]: the run-cutting heuristic, the matching algorithm and the set covering formulation. The run-cutting algorithm was used in the 1970's, this is a constructive algorithm used by manual schedulers. The matching method is divided into three parts: block partition of the timetable, graph generation and duty achievement. Although in the early eighties several researchers recognized the need to integrate vehicle and crew scheduling, most of the algorithms published in the literature still follow the sequential approach where vehicles are scheduled before, and independently of, crews. In the operations research literature, only a few applications address a simultaneous approach to vehicle and crew scheduling. None of those publications makes a comparison between simultaneous and sequential scheduling [2].

II. LITERATURE ON LOCOMOTIVE ASSIGNMENT AND CREW SCHEDULING

Booler [3] proposed a Lagrangean relaxation method for solving an integer programming model of the locomotive assignment problem. Cordeau et al. [5] presented a survey of recent optimization models for the most commonly studied rail transportation problems. They described the problem of simultaneous assignment of locomotives and cars to passenger trains. An exact algorithm, based on Bender's decomposition, is proposed to solve the problem. Kron and Fschetti [6] described the intelligent information systems that are used by Dutch railway operator for supporting the scheduling of drivers and guards. Their model is a set covering model and they solved the model by applying dynamic column generation techniques, Lagrangean relaxation and powerful heuristics.

Cordeau et al. [7] proposed a heuristic solution approach based on mathematical optimization for the assignment of locomotives and cars to scheduled trains. Ernst et al. [8] presented a model for integrated optimization model for train crew management. They used mathematical programming model for solving the problem. Walker et al [9] prepared the train time table and crew roster at the same time. They solved the problem in two phases. In the first phase they prepare train time tale, and in the second phase according to train time table they prepared crew roster. They used branch and bound algorithm for solving two phases.

Rouillon et al. [10] addressed the operational locomotive assignment problem of providing sufficient motive power to pull a set of scheduled trains at minimum cost while satisfying locomotive availability and maintenance requirement. Godwin et al [11] solved the locomotive assignment problem in two phases. In the first phase, they assigned locomotives with partial scheduling with the objectives of minimizing total deadheading time and total coupling delay. They used a genetic algorithm to find non-dominant locomotive assignment solutions and proposed a method for evaluating its performance. The solutions are then ranked using two approaches, based on the decision maker's preferences. In the second phase, they selected a locomotive assignment solution based on the ranking and find the lower bound on the arrival time of freight trains at their destinations. They used a genetic algorithm again to schedule the freight trains in the passenger rail network, with prescribed locomotive assignment precedence constraints with the objective of minimizing total tardiness. Ghoseiri and Morshedsolouk [12] developed an algorithm for the train scheduling problem using the ant colony system meta-heuristic called ACS-TS. The problem is considered as a traveling salesman problem (TSP) where in cities represent the trains.

III. Proposed Approach

Crew scheduling is characterized as the construction of duties in such a way that the timetable is covered adequately [8]. The aim of this study is to integrate locomotive assignment and crew scheduling problem by using mixed integer linear programming approach. Decisions about repositioning engines and crews made mainly on the basis of expert judgment. If extra engines and crews cannot be carried by regularly scheduled trains, engines are dispatched in light moves from the nearest yard and crews are moved from one yard to another without an engine. The planning horizon for crew scheduling is one week.

The definitions of the key parameters and decision variables of the proposed model can be developed as follows:

Indices

j, k	Set of yards
i	Set of days

Parameters

J	Number of yards under consideration
F_{jk}	Fixed cost of moving light engines from $Yard_j$ to $Yard_k$ in dollars
V_{jk}	Variable cost of moving one light engine from $Yard_j$ to $Yard_k$ in dollars per engine
D_{ij}	Number of engines required on Day_i at $Yard_j$
G_{ij}	Number of engines gained or lost on Day_i at $Yard_j$
S	Size of locomotive fleet
L	Size of crew fleet
C_{jk}	Variable cost of delivering a crew from $Yard_j$ to $Yard_k$ in dollars without an engine
T_{ij}	Number of crews required on Day_i at $Yard_j$
K_{ij}	Number of crews gained or lost on Day_i at $Yard_j$

Decision variables

b_{ijk}	$\begin{cases} 1 & \text{if a light engine move takes place on Day}_i \text{ from } Yard_j \text{ to } Yard_k \text{ (} k \neq j \text{)} \\ 0 & \text{otherwise} \end{cases}$
x_{ijk}	Number of light engines moved on Day_i from $Yard_j$ to $Yard_k$
y_{ij}	Number of engines available at the beginning of Day_i at $Yard_j$
z_{ijk}	Number of crews moved on Day_i from $Yard_j$ to $Yard_k$ with an engine
p_{ijk}	Number of crews moved on Day_i from $Yard_j$ to $Yard_k$ without an engine
m_{ij}	Number of crews available at the beginning of Day_i at $Yard_j$

The integrated locomotive utilization and crew scheduling problem may be formulated as the following mixed integer linear program:

$$\text{Minimize } Z = \sum_{i=1}^J \sum_{j=1}^J \sum_{\substack{k=1 \\ k \neq j}}^J [(F_{jk} b_{ijk} + V_{jk} x_{ijk}) + p_{ijk} C_{jk}] \quad (1)$$

Subject to

$$y_{ij} \geq D_{ij}, \quad 1 \leq i \leq 7, \quad 1 \leq j \leq J \quad (2)$$

$$\sum_{\substack{n=1 \\ n \neq j}}^J x_{ijn} \leq y_{ij} + G_{ij}, \quad 1 \leq i \leq 7, \quad 1 \leq j \leq J \quad (3)$$

$$y_{ij} + G_{ij} + \sum_{m=1}^J x_{imj} - \sum_{n=1}^J x_{ijn} = y_{i+1,j}, \quad 1 \leq i \leq 6, \quad 1 \leq j \leq J \quad (4a)$$

$$y_{7j} + G_{7j} + \sum_{m=1}^J x_{7mj} - \sum_{n=1}^J x_{7jn} = y_{1,j}, \quad 1 \leq j \leq J \quad (4b)$$

$$x_{ijk} \leq 15 b_{ijk}, \quad 1 \leq i \leq 7, \quad 1 \leq j \leq J, \quad 1 \leq k \neq j \leq J \quad (5)$$

$$\sum_{j=1}^J y_{ij} \leq S, \quad 1 \leq i \leq 7 \quad (6)$$

$$b_{ijk} \in \{0,1\}, \quad 1 \leq i \leq 7, \quad 1 \leq j \leq J, \quad 1 \leq k \neq j \leq J \quad (7)$$

$$x_{ijk} \text{ and } y_{ij} \text{ integer}, \quad 1 \leq i \leq 7, \quad 1 \leq j \leq J, \quad 1 \leq k \neq j \leq J \quad (8)$$

$$m_{ij} \geq T_{ij}, \quad 1 \leq i \leq 7, \quad 1 \leq j \leq J \quad (9)$$

$$\sum_{\substack{n=1 \\ n \neq j}}^J z_{ijn} + \sum_{\substack{n=1 \\ n \neq j}}^J p_{ijn} \leq m_{ij} + K_{ij}, \quad 1 \leq i \leq 7, \quad 1 \leq j \leq J \quad (10)$$

$$m_{ij} + K_{ij} + \sum_{\substack{m=1 \\ m \neq j}}^J z_{imj} + \sum_{\substack{m=1 \\ m \neq j}}^J p_{imj} - \sum_{\substack{n=1 \\ n \neq j}}^J z_{ijn} - \sum_{\substack{n=1 \\ n \neq j}}^J p_{ijn} = m_{i+1,j}, \quad 1 \leq i \leq 6, \quad 1 \leq j \leq J \quad (11a)$$

$$m_{7j} + K_{7j} + \sum_{\substack{m=1 \\ m \neq j}}^J z_{7mj} + \sum_{\substack{m=1 \\ m \neq j}}^J p_{7mj} - \sum_{\substack{n=1 \\ n \neq j}}^J z_{7jn} - \sum_{\substack{n=1 \\ n \neq j}}^J p_{7jn} = m_{1,j}, \quad 1 \leq j \leq J \quad (11b)$$

$$z_{ijk} \leq L b_{ijk}, \quad 1 \leq i \leq 7, \quad 1 \leq j \leq J, \quad 1 \leq k \neq j \leq J \quad (12)$$

$$\sum_{j=1}^J m_{ij} \leq L, \quad 1 \leq i \leq 7 \quad (13)$$

$$z_{ijk}, p_{ijk} \text{ and } y_{ij} \text{ integer}, \quad 1 \leq i \leq 7, \quad 1 \leq j \leq J, \quad 1 \leq k \neq j \leq J \quad (14)$$

The objective function to be minimized includes fixed and variable costs for engine movements and also variable cost for crew transfers without an engine. The constraints (2,3,...,8) are related with the engine utilization and the others are related with crew allocation. Constraints 2 used to satisfy the number of engines available at each yard at the start of each day to be large enough to haul the scheduled outgoing trains. In Constraints 3, the number of light engines moved out of each yard on each day is limited to the sum of the engine availability at the beginning of the day at the yard and the net gain or loss of engines determined by the incoming and outgoing trains. Constraints 4a and 4b are ensured that, at each yard, the total number of engines available at the beginning of each day is equal to the initial availability plus the net gain or loss of engines due to various light engine moves in the previous day. In particular, constraints 4b are included to bridge the gap between the Sunday of one week and the Monday of the next so that the cyclical nature of the engine flows are preserved. Constraints 5 stated that the number of engines moved in any deadheading trip on any day must be smaller than or equal to 15, which is an operating rule. Constraints 6 specified that, at the beginning of each day, the total number of engines available at the various yards do not exceed the current locomotive fleet size. While this inequality involves seven distinct constraints, one for each day of the week, it can be easily proven that only one of them is needed due to the flow conservation of engines within the closed railroad system as required by constraints 4a and 4b. Finally, in Constraints 7, b_{ijk} is defined as a binary variable whereas both x_{ijk} and y_{ij} are to take on only non-negative integer values in Constraints 8.

Constraints 9 provide the required crews for each yard at the start of each day to be able to schedule necessary crew for outgoing trains. The number of crews moved out of each yard on each day with or without

engine is limited to the sum of the crews available at the beginning of the day at the yard and the net gain or loss of crews determined by the incoming and outgoing crews are declared in Constraints 10. Constraints 11a and 11b are guaranteed that at each yard, the total number of crews available at the beginning of each day is equal to the initial availability plus the net gain or loss of crews due to various crew deliveries in the previous day. Constraints 12 assure that if there is a move from $Yard_i$ to $Yard_j$ then all of the required crews can be delivered by that engine, if not the required crews should be delivered with another way by paying variable transport costs. The total number of crews available at the various yards does not exceed the total crew size, at the beginning of each day is ensured in Constraints 13. Constraints 14 include the binary and integer restrictions.

IV. AN ILLUSTRATIVE APPLICATION

The data sets including the crew and engine requirements at $Yard_j$ ($j=1,2,3$) and fixed costs and variable cost are determined by expert judgment. The data sets that include engine and crew requirements for each yard for each day are displayed in Appendix A. Also, the variable and fixed costs for moving engines and delivering crew without an engine are shown in Appendix B.

Solving the mathematical program using LINDO, the following solutions are observed: $(b_{112}^*, b_{332}^*, b_{612}^*, b_{632}^*, b_{712}^*, b_{723}^*)=(1,1,1,1,1,1)$, $(x_{112}^*, x_{332}^*, x_{612}^*, x_{632}^*, x_{712}^*, x_{723}^*)=(3,2,2,2,4,1)$, $(y_{11}^*, y_{12}^*, y_{13}^*, y_{21}^*, y_{22}^*, y_{23}^*, y_{31}^*, y_{32}^*, y_{33}^*, y_{41}^*, y_{42}^*, y_{43}^*, y_{51}^*, y_{52}^*, y_{53}^*, y_{61}^*, y_{62}^*, y_{63}^*, y_{71}^*, y_{72}^*, y_{73}^*)=(6,7,3,5,7,4,7,6,3,6,7,3,8,6,2,7,7,2,5,9,2)$, $(p_{421}^*, p_{521}^*)=(6,3)$, $(z_{332}^*, z_{632}^*, z_{712}^*)=(16,2,5)$, $(m_{11}^*, m_{12}^*, m_{13}^*, m_{21}^*, m_{22}^*, m_{23}^*, m_{31}^*, m_{32}^*, m_{33}^*, m_{41}^*, m_{42}^*, m_{43}^*, m_{51}^*, m_{52}^*, m_{53}^*, m_{61}^*, m_{62}^*, m_{63}^*, m_{71}^*, m_{72}^*, m_{73}^*)=(9,8,8,6,8,10,7,5,18,5,19,6,6,14,4,8,13,3,8,14,3)$ which means that, light engine should be moved from $Yard_1$ to $Yard_2$ on Monday, from $Yard_3$ to $Yard_2$ on Wednesday, from $Yard_1$ to $Yard_2$ on Saturday; and so on. Three light engines should be moved from $Yard_1$ to $Yard_2$ on Monday, 4 light engines should be moved from $Yard_1$ to $Yard_2$ on Sunday; and so on. There should be six engines at $Yard_1$, seven engines at $Yard_2$ on Monday, five engines at $Yard_1$, and four engines at $Yard_3$ on Tuesday. Six crews should be moved without an engine from $Yard_2$ to $Yard_1$ on Thursday, three crews should be moved without an engine from $Yard_2$ to $Yard_1$ on Friday. 16 crews should be moved with engine from $Yard_3$ to $Yard_2$ on Wednesday, two crews should be moved with engine from $Yard_3$ to $Yard_2$ on Saturday, five crews should be moved with engine from $Yard_1$ to $Yard_2$ on Sunday. There should be nine crews at $Yard_1$, eight crews at $Yard_2$, eight crews at $Yard_3$ on Monday, seven crews at $Yard_1$, five crews at $Yard_2$ on Wednesday. The objective function value is $z^*=4938,22$.

V. CONCLUSIONS

Crew scheduling is a well-known problem in Operations Research, and until recently has been associated with many real life examples. The crew scheduling problem for railways has been a major topic for researchers and increased computing power allows solving of crew scheduling problems optimally even for large instances.

To our knowledge there is not adequate number of study in literature, integrating crew scheduling problem with locomotive assignment. In this paper integrated locomotive assignment and crew scheduling problem is proposed by using mathematical linear programming model and solved optimally with LINDO in very short time. Optimization model is validated by a numerical example and achieved results show that integrated crew scheduling and locomotive assignment model is an effective tool. Future research can be studying the integrated model with larger number of instances and also developing heuristic algorithm in real-life applications.

APPENDIX A

Engine/crew requirements and availabilities are shown below:

		Engine Requirements			Total Outbound	Net gain/loss
Day of Week	From \to	Yard 1	Yard 2	Yard 3		
Monday	Yard 1	0	4	2	6	2
	Yard 2	5	0	2	7	-3
	Yard 3	3	0	0	3	1
	Inbound	8	4	4		
Tuesday	Yard 1	0	4	1	5	2
	Yard 2	5	0	0	5	-1
	Yard 3	2	0	0	2	-1
	Inbound	7	4	1		
Wednesday	Yard 1	0	3	2	5	-1
	Yard 2	2	0	2	4	-1
	Yard 3	2	0	0	2	2
	Inbound	4	3	4		
Thursday	Yard 1	0	3	1	4	2
	Yard 2	4	0	0	4	-1
	Yard 3	2	0	0	2	-1
	Inbound	6	3	1		
Friday	Yard 1	0	4	1	5	-1
	Yard 2	3	0	0	3	1
	Yard 3	1	0	0	1	0
Saturday	Inbound	4	4	1		
	Yard 1	0	5	2	7	0
	Yard 2	5	0	2	7	-2
	Yard 3	2	0	0	2	2
	Inbound	7	5	4		
Sunday	Yard 1	0	4	1	5	5
	Yard 2	8	0	1	9	-5
	Yard 3	2	0	0	2	0
	Inbound	10	4	2	16	

Crew Requirements						
Day of Week	From \to	Yard 1	Yard 2	Yard 3	Total Outbound	Net gain/loss
Monday	Yard 1	0	8	3	11	-3
	Yard 2	5	0	3	8	0
	Yard 3	4	0	0	4	2
	Inbound	9	8	6		
Tuesday	Yard 1	0	4	8	12	-5
	Yard 2	5	0	2	7	-3
	Yard 3	2	0	0	2	8
	Inbound	7	4	10		
Wednesday	Yard 1	0	3	3	6	-2
	Yard 2	2	0	3	5	-2
	Yard 3	2	0	0	2	4
	Inbound	4	3	6		
Thursday	Yard 1	0	3	2	5	1
	Yard 2	4	0	0	4	1
	Yard 3	2	2	0	4	-2
	Inbound	6	5	2		
Friday	Yard 1	0	4	2	6	-1
	Yard 2	3	0	0	3	2
	Yard 3	2	1	0	3	-1
	Inbound	5	5	2		
Saturday	Yard 1	0	6	2	8	0
	Yard 2	5	0	2	7	-1
	Yard 3	3	0	0	3	1
	Inbound	8	6	4		
Sunday	Yard 1	0	4	1	5	6
	Yard 2	8	0	7	15	-11
	Yard 3	3	0	0	3	5
	Inbound	11	4	8	23	

APPENDIX B

Light engine move costs and variable cost per crew transfer without engine are shown below.

Light engine move costs

From \to	Yard 1	Yard 2	Yard 3
<i>Fixed cost per move</i>			
Yard 1		463,84	1284,96
Yard 2	463,84		896,34
Yard 3	1284,96	896,34	
<i>Variable cost per engine</i>			
Yard 1		37,57	135,33
Yard 2	37,57		100,31
Yard 3	135,33	100,31	
<i>Variable cost per crew transfer without engine</i>			
Yard 1		2	4
Yard 2	2		3
Yard 3	4	3	

REFERENCES

[1] L.D. Bodin, B.L. Golden, A.A. Assad, and M.O. Ball. Routing and scheduling of vehicles and crews. Computers & Operations Research, 10(63):2-12, 1993.

[2] R. Freeling, D. Huisman, and A. Wagelmans. Models and algorithms for integration of vehicle and crew scheduling. Economic Institute Report, EI2000-10/A.

[3] J.M.P. Boaler. A note on the use of Lagrangean relaxation in railway scheduling. Journal of the Operational Research Society, 46: 123-127, 1995.

[4] J.F. Cordeau, P. Toth, and D. Vigo. A survey of optimization models for train routing and scheduling. Transportation Science, 32 (4): 380-404, 1998.

[5] J.F. Cordeau, F. Soumis, and J. Desrosiers. A Benders decomposition approach for the locomotive and car assignment problem. Transportation Science, 34 (2): 133-149, 2000.

[6] L. Kroon, and M. Fischetti. Scheduling train drivers and guards: the Dutch “Noord-Oost” Case”. Proceedings of the 33rd Hawaii International Conference on System Sciences, 2000.

[7] J.F. Cordeau, G. Desaulniers, N. Lingaya, F. Soumis, and J. Desrosiers. Simultaneous locomotive and car assignment at VIA Rail Canada. Transportation Research – B, 35: 767-787, 2001.

[8] A.T. Ernst, H. Jiang, M. Krishnamoorthy, H. Nott, and D. Sier. An integrated optimization model for train crew management. Annals of Operations Research, 108: 211-224, 2001.

[9] C.G. Walker, J.N. Snowdon, and D.M. Ryan. Simultaneous disruption recovery of a train timetable and crew roster in real time. Computers & Operations Research, 32: 2077-2094, 2005.

[10] S. Rouillon, G. Desaulniers, and F. Soumis. An extended branch and bound method for locomotive assignment. Transportation Research Part B, 40(5): 404-423, 2006.

[11] T. Godwin, R. Gopalan, and T.T. Narendran. Locomotive assignment and freight train scheduling using genetic algorithms. International Transactions in Operational Research, 13: 299-332, 2006.

[12] K.F. Ghoseiri, and F. Morshedsolouk. ACS-TS: Train scheduling using ant colony system. Journal of Applied Mathematics and Decision Sciences, 2006: 1-28, 2006.

Search Engine Using Spatial Data

P.Sreedevi¹, G.Sridevi², B.Padmaja³

M.Tech. (CSE), NCET¹, Associate Professor, Department of CSE², Assistant Professor, DRKCET³

ABSTRACT:

Spatial search engines are specialized search engines primarily dedicated to retrieve geographical information through web technology. They provide capabilities to query metadata records for related spatial data, and link directly to the online content of spatial data themselves. OpenSearch-Geo extensions are developed to facilitate basic geographical data search using Open-search method.

OpenSearch-Geo extensions add new parameters of geographic filtering for querying spatial data and recommended set of simple standards responses in geographic format, such as KML, Atom and GeoRSS through spatial search engines. The communication method used in spatial search engines is based on standardized- Service-Oriented –Architecture. In this catalogue service plays a significant role. It provides a common mechanism to classify, register, describe, search, maintain and access information about resources available on a network.

In the contemporary search engine there is no mechanism to find out which of the available resources are best fit for use to users. We propose search functionality for current spatial data search engines to consider user quality requirements in addition to the geographical extent and keyword matching

KEYWORD: Spatial, KML, Atom and GeoRSS

I. INTRODUCTION:

Usage of spatial data resources on the web has increasingly become important in daily activities of modern society. In web technology mainstream search engines like Google, Yahoo, and ALO are used in accessing distributed information. When Internet users type a keyword or phrase into the search engine query box, they expect a list of search results which can be websites that offer information, products or services related to that keyword. However, finding proper web content is difficult due to availability of vast volume of information on the web. Therefore, searching for proper result requires specialized search engines.

BACKGROUND AND RELATED WORK

The motivation of this paper is to develop method for reasoning based on fitness for use to enable spatial search engines recommending spatial data resource for users.

Presently, there is no spatial data search engine that reason out based on consideration of fitness for use. We performed extensive study on the concepts in fitness for use and recommendation technologies. We also studied the quality of spatial data and users quality requirements to determine fitness for use.

We reviewed literature on fitness for use approach from users and producers perspective in spatial data infrastructure (SDI). We also studied the quality of spatial data and users quality requirements to determine fitness for use.

We designed the profiling algorithm and a reasoning logic to determine fitness for use of spatial data resources. We implement the model and the reasoning logic by using UML modelling language using the Enterprise Architect. For the realization of our system, we use the PostgreSQL spatial database management system with PHP: Hypertext Preprocessor for developing the front end as a web application system. The fitness for use reasoning logic is implemented by using the PostgreSQL structured language (PL/pgSQL) database programming languages.

FITNESS FOR USE AND RECOMMENDATION SYSTEMS

Fitness for use is a way of understanding the relationship between data and the users. The definition of fitness for use has been subjected to the usability of datasets. Redman suggested that for dataset to be fit for use it must be accessible, accurate, timely, complete, consistent with other sources, relevant, comprehensive, provide a proper level of detail, be easy to read and easy to interpret. Therefore, fitness for use can be viewed as the capability of the dataset to fit stated user requirements and application specifications.

1.1 DATA QUALITY VERSUS FITNESS FOR USE

Data quality is a perception or an assessment of data's fitness to serve its purpose in a given context and subjective to various applications. It highly depends on the need of individuals on how to use datasets. Quality can be described by object or phenomenon attributes and properties.

The term data quality is used to describe the correspondence between an object in reality and its representation in the datasets. Quality can also be expressed as a measure against a production specification or user requirements.

In GIS context a quality product is a product which is free from errors, or a product with confirmation of specifications used, or it can be a product that satisfies user's expectations. However, widely accepted expression affirms that spatial data quality is recognized only in terms of its specific use.

Organization ISO is accepted in common to describe spatial data quality. The ISO defines quality as the totality of characteristics of a product that bear on its ability to satisfy stated and implied needs. Therefore, for ISO quality is a result that has to be observed during use.

The standards mainly describe the spatial data quality using two main categories: quality overview elements and quantitative quality elements. The ISO provide quality elements with their sub-elements and guidelines for producers to describe the characteristics of the datasets. Spatial data quality evaluation procedure and reporting the result for quality evaluation procedure are also defined by ISO standards.

Devillers and Jeansoulin elaborates the concept of spatial quality by dividing it into two: internal and external quality. Internal quality is used to express the products with no errors. The internal quality describes characteristics that define the apparent individual nature of products. On the other hand, external quality is used to express products that meet user needs. It is associated to express the similarity between the data produced and user requirements and their needs.

When data quality description is defined by fitness for use, it should assure the user that the datasets are fit for the intended use.

1.2 FITNESS FOR USE: SPATIAL DATA PRODUCERS' PERSPECTIVE & USERS' PERSPECTIVE

In Geographic Information Science (GIS) environments spatial datasets frequently have different origins and contain different quality levels.

Producers' perception of spatial data quality mainly depends on the dataset's internal characteristics. These intrinsic characteristics are resulted from production methods, e.g. data acquisition technologies, data models, and storages. Internal quality description of spatial dataset is independent of any task, unless it is collected and processed for a specific application. Producers of spatial data resource assume that users are able to determining a spatial dataset's fitness for use before use of the dataset. Under the fitness for use approach, producers do not make any judgment. Spatial data producers provide quality information contents is to help users to determine if spatial datasets fulfill their application's quality requirements.

Spatial data user's quality requirements are rooted in the intended application they want the dataset to be used for. Users usually evaluate fitness for use of data sources to determine the suitability of data for problem solving and decision making and consider the datasets interoperability with other data sources. In addition, users also determine fitness for use according to their multidisciplinary information needs. Other factors, such as compliance to specific needs and availability of rules and quality control also have impact for users to determine fitness for use.

Directly or indirectly users of a dataset need to use information about spatial data quality in order to be able to assess the fitness for use of the data in their context.

1.3 APPROACHES TO DETERMINE FITNESS FOR USE

Determining fitness for use of a data resource is the only method to avoid risks caused by misuse of spatial data. Comprehensive comparison against user quality requirement and detailed quality description of dataset is the main approach to determine fitness for use. In determining fitness for use user's quality requirements, quality description of the dataset, the decision and how it will be influenced by quality are required input parameters. Given these information, evaluation of fitness for use can be implemented. For fitness for use evaluation, the user quality requirement and the dataset quality requirement should have the same base point. Otherwise, with the absence of such common agreement on quality of object, fitness for use assessment becomes much more complicated.

Each user group has certain requirements and different aspects of usability that have to be considered. The fitness for use decision can be easily determined if users quality requirement is known.

The well known approach in understanding user's quality requirements is translating subjective user's requirements into an objective technical specification. As a general approach in this research the reasoning logic design to determine fitness for use of spatial dataset is also based on comparison of user quality requirement (external quality) and the dataset quality description (internal quality).

1.4 RECOMMENDER SYSTEMS

Recommender systems are widely implemented for searching, sorting, classifying, filtering and sharing a vast amount of information available on the web to allow users to find resources that fit their need. All recommender systems take advantage of a particular set of artificial intelligence techniques. Recommender systems represent user preferences for the purpose of suggesting items to the users so that users are directed toward those items that best meet their needs and preferences. A recommender system customizes its responses to a particular user. Instead of direct response to queries, a recommender system is intended to serve as an information agent of individual users or group of users.

Recommendation techniques have a number of possible classifications. However, all recommender systems have three common fundamental components. The first component referred to as background data is the information that the system had before the recommendation process begins. The second component is the information that users must communicate to the system in order to generate a recommendation. It is referred to as input data. The third is the algorithm that combines background information and input data.

Recommendation techniques can be distinguished on the basis of their knowledge sources which can be the knowledge of other users' preferences, ontological or inferential knowledge about the domain, or added by users themselves. The main classification of recommendation techniques are:

- **Collaborative filtering:** Collaborative recommendation is probably the most familiar, most widely implemented and most mature among existing recommendation technologies. Collaborative recommender systems aggregate ratings or recommendations of objects, recognize commonalities between users on the basis of their ratings, and generate new recommendations based on inter-user comparisons. Griffith et.al conducted a survey on performance of collaborative filtering.

- **Content-based:** The system generates recommendations from two sources: the features associated with products and the ratings that a user has given them. Content-based recommender systems treat recommendation as a user-specific classification problem and learn a classifier for the user's likes and dislikes based on product features. A content-based recommender learns a profile of the user's interests based on the features present in objects the user has rated. It is item-to-item or user-to-user correlation. Decision trees, neural nets, and vector-based representations have all been used. As in the collaborative case, content-based user profiles are long term models and updated as more evidence about user preferences is observed.

- **Hybrid recommender systems** combine two or more recommendation techniques to gain better performance with fewer of the drawbacks of any individual one. Most commonly, collaborative filtering is combined with some other technique in an attempt to avoid the ramp-up problem.

Recommender systems typically determine matches via a process of identifying similar users by creating neighbor users. Determining recommendations based on selected neighbors is named as profile matching. Profile matching involves:

- Find similar users: employing standard similarity measures technique such as Nearest neighbour, Clustering and Classification
- Create a neighbour: techniques used include the creation of centroid, correlation-thresholding, and best-n-neighbours.

PROFILE BUILDING AND MAINTENANCE

The generation and maintenance of accurate user profiles is an essential component of a successful recommender system. Consequently, in analyzing how a recommendation system makes individuals recommendations or assesses a user needs, the key issue is the user profile. A recommender agent cannot begin to function until the user profile has been created.

RECOMMENDATION SYSTEM DATA MODEL DESIGN AND REASONING LOGIC

1.5 INTRODUCTION

In Section 2, we discussed the concepts of fitness for use from users and producers perspective. Both spatial data users and producers agree that fitness for use evaluation of a dataset before its usage reduce risks caused by misuse of spatial data resource. However, the two sides are not in line with the definition of fitness for use. The concept of fitness for use for spatial data users is the dataset that satisfies their need based on their quality requirement. On the other hand producers express fitness for use as the description of quality description of the dataset. Hence, the assessment and determination of fitness for use of a dataset remain users' responsibility. However fitness for use computation is not an easy task for users.

Adapting such an approach to the spatial data search engine is of great importance to search spatial data based on fitness for use. In this research work we propose a mechanize to store users spatial data search quality requirements and spatial data quality descriptions that can be used in fitness for use evaluation to recommend spatial datasets to users based on their requirements.

1.6 SPATIAL DATA RECOMMENDATION SYSTEM ARCHITECTURE

The proposed recommendation system design involves three main components as shown in figure. The figure describes the general view of spatial data recommendation system design framework.

- User interface: allows and controls user system interaction. The recommendation service obtains information about users' need through web based user interface. The user interface design considers user groups. For example, expert users group requires detailed quality information to determine if the resource is useful for their task or not. However, non GIS expert users group lacks understanding about detailed quality information. Therefore, the user interface design should support simple way of allowing these users to specify their data quality requirement. Moreover, if the users group is non human users, special web service communication facility like XML/GML standard data format should be maintained.

- Recommendation system: is the main component of the system which controls the overall interaction to provide fitness for use based spatial data recommendation.

- Profile database: is the data model of the recommendation system which store users information and spatial data quality information in a structured form. It allows automatic and active data retrieval to speed up the fitness for use evaluation, prediction and recommendation process of spatial data resources. Structured profile storage is defined by the conceptual data model of spatial data recommendation system which is discussed in the following section in detail.

FITNESS FOR USE EVALUATION FUNCTIONALITY

After users spatial data search requirements and spatial data resources information are profiled in the spatial data recommendation data model, in order to recommend the spatial data resources for users, the system should make fitness for use evaluation. In this research we discuss the fitness for use evaluation from three aspect: spatial extent matching, application matching with spatial data resources description and overview quality elements, and quantitative data quality evaluation aspect. However, since the fitness for use evaluation is performed using the system data model, the sequence does not have difference in recommending the datasets for users.

FITNESS FOR USE EVALUATION USING SPATIAL EXTENT

Fitness for use evaluation using user spatial extent requirement requires extensive spatial matching to get dataset with the best fit extent. First of all the spatial data resources that have spatial extent matching with users spatial extent requirement needs to be filtered. All the datasets which have intersection with user spatial extent requirements will be returned as a candidate dataset for further filtering.

This phase of filtering spatial datasets needs to be addressed from different aspect of spatial extent matching functions. For example, the user extent requirement may be completely inside the dataset extent or only a portion of area of user extent may intersect with the dataset extent. Therefore, spatial area difference can be known by calculating the area ratio. Hence, area ratio computation of intersection with user spatial extent requirement and area ratio computation of intersection with spatial data resources extent helps to identify the best fit spatial data resources. The value of area ratio is given in percentage.

Then by sorting datasets descending using the ratio of intersection and user extent the system can identify and return the best datasets. If there are more datasets that have similar area ratio values, again the ratio of intersection and dataset extent help us to identify the best one. Based on this logic we design algorithm 6 to rank spatial datasets using the computed area ratio.

In the algorithm design for the spatial computation the following PostGIS built in functions are used:

- *ST_GeometryType* Return the geometry type of the ST_Geometry value.
- *ST_within*: returns true if one geometry is within the geometry of the other, it takes two arguments. We used the user extent and spatial dataset extent to return true or false
- *ST_Centroid* : This function takes one argument. We used it to return the centroid of the geometry given by the user as a point. Therefore, the centre of the user extent requirement can be check within the extent of dataset.
- *ST_Intersects* : generates a boolean result after checking intersection between two geometry
- *ST_Intersection* : takes two ST_Geometry objects and returns the intersection set as an ST_Geometry object.
- *ST_GeomFromText* : returns a specified ST_Geometry to be enable the spatial function work

Variable definition used in Algorithm 1 - 3:

- UA - user application
- $DS_i \quad i=1 \dots N \in DSS$ - where DSS is set of selected datasets
- $DS_j \quad j=1 \dots N \in DS \quad S$ where DS is set of datasets E
- I_E - user and dataset intersection extent E
- AI - UE and DSE intersection area
- RI_U - AI and area of UE ratio
- RI_{DS} - AI and area of DSE ratio
- DS_iSE - selected extent dataset
- DSSS - selected and sorted dataset
- DSS - sorted DSSS

Algorithm 1: Select dataset based on user extent

Procedure:

- for all datasets, select a dataset if:
- user extent is within dataset extent
- center of user extent is within the dataset extent
- user extent and dataset extent intersection has polygon geometry
- return DS_S

Input: DS, U_E, DS_E

- 1: for DS^i to DS^N do
- 2: $DS^{iE} \leftarrow \text{extract_dataset_extent}(DS^i)$
- 3: if $ST_within(U_E, DS^{iE})$ then
- 4: $DS_S \leftarrow DS^i$

```

5: else if ST_within(ST_centroid( $U_E$ ,  $DS^iE$ )) then
6:  $DS_S \leftarrow DS^i$ 
7: else
8: if ST_Intersect( $U_E$ ,  $DS^iE$ ) then
9:  $I^i \leftarrow ST\_Intersection(U_E, DS^iE)E$ 
10: if ST_GeometryType(ST_GeomFromText( $A_I$ )) = "ST_Polygon" then
11:  $DS_S \leftarrow DS^i$ 
12: end if
13: end if
14: end if
15: end for
16: return  $DS_S$ 

```

Spatial datasets which have a polygon intersection with user spatial extent requirement are returned as a result of algorithm 1. Once the spatial datasets are filtered by the spatial extent matching as given by algorithm 1, the selected datasets will be an input for algorithm 2.

Algorithm 2: Area ratio computation

Procedure:

- for all selected datasets:
- calculate user extent and dataset extent intersection
- compute intersection and user extent area ratio
- compute intersection and dataset extent area ratio
- return dataset selected, area ratios

Input: DS_S , U_E , DS^iE

```

for  $DS^iS$  to  $DS^N$  do S
 $I^i \leftarrow ST\_Intersection(U_E, DS^iSE)E$ 
 $R^i_{-U} \leftarrow \frac{ST\_Area(I^i)}{ST\_Area(U_E)}$ 
 $R^i_{-D} \leftarrow \frac{ST\_Area(I^i)}{ST\_Area(DS^iE)}$ 
end for
return  $DS_S$ ,  $R_{-U}$ ,  $R_{-D}$ 

```

Algorithm 2 returns the same dataset that has been returned by algorithm 1 with newly computed area ratio extent information. This area ratio helps to order the dataset in order to identify the best one. The ranking procedure using spatial extent ratio value for every candidate spatial dataset is given in algorithm 3. However, for simplicity purpose we use sort function supported in PostGIS for implementation as given in algorithm 4.

Algorithm 3: Rank dataset based on Extent ratio

Procedure:

- sort selected dataset by R_{-U} (A_I and area of U_E ratio)
- for all selected and sorted datasets, if two or more consecutive datasets have equal R_{-U} , sort these rows with R_{-D} (A_I and area of DS^iE ratio) else update the index to indicate to the next group

Input: DS_S , R_{-U} , R_{-D}

```

1:  $DS_S \leftarrow \text{sort } DS_S \text{ desc } R_{-U}$ 
2: for  $i$  to  $M$  do
3: //where  $M$  is the number of selected and sorted datasets
4: if  $R^i_{-U} = R^{i+1}_{-U}$  then  $_U$ 
5: for  $j = i$  to  $M$  do
6: if  $R^i_{-D} = R^j_{-D}$  then

```

```

7: temp  $\square \leftarrow DS^j SS$ 
{temporarily save current record  $DS^j$  }  $SS$ 
8: //next two lines swap current record with the next
9:  $DS^j \square_{SS} \leftarrow DS^{j+1} SS$ 
10:  $DS^{j+1} \square \leftarrow temp SS$ 
11: end if
12: //to check the next group having the same  $R_{L_U}$ 
13: if  $R_{L_U}^{j+1} = R_{L_U}^{j+2}$  then
14: //if the next two records  $A_I$  and area of  $U_E$  ratio are not equal,
15: //set position for the next comparison to this group and break the
16: //inner loop
17:  $i=j+2$ 
18: break
19: end if
20: end for
21: end if
22: end for
23: return  $DS^S SS$ 

```

Algorithm 4: Rank dataset based on Extent ratio

Procedure:

- sort selected dataset by R_{L_U} (A_I and area of U_E ratio)
 - for all selected and sorted datasets, if two or more consecutive datasets have equal R_{L_U} , sort these rows with R_{L_D} (A_I and area of DS_E ratio)
- return $DS^S SS$

```

1:  $DS^S \square \leftarrow \text{sort } DS_S \text{ desc } R_{L_U}, \text{ desc } R_{L_D} SS$ 
2: return  $DS^S SS$ 

```

FITNESS FOR USE EVALUATION USING APPLICATION

To design the fitness for use evaluation based on user application requirement, it is required to define theme_keyword that represent the application by referring thematic classification of dataset. The concept of theme_keyword definition is mainly required to search different spatial datasets which can be useful for the intended application. The theme_keyword definition for the application gives wide range of possibly to search various resources for the intended application. The ISO standard metadata representation topic categories is one of metadata elements required to identify a dataset, that is used to group keywords and to learn more about main themes of the dataset to understand topics exist in the dataset description. It is high-level geographic data thematic classification to assist in the grouping and search of available geographic datasets. The topic category is also used for topic-based search of available spatial data resources. It is one of a handful element that describes the type of features that are included in a dataset.

Variable definition used in Algorithm 5-9:

- U_A - user application
- $DS^i S \square_{i=1 \dots N} \in DS_S$ - where DS_S is set of selected datasets based on U_A or TKW
- $TKW_j \square_{j=1 \dots M} \in TKW$ - where TKW is set of theme_keywords
- $\Omega_j^T KW \square \in \square \square, 1 \square$ - weight assigned for theme_keyword j
- Ω_A - weight assigned for U_A
- DS - dataset
- O_Q - overview quality description of DS^i
- $N = \square \square \mathcal{B} \square$ - number of datasets DS in database
- M - number of theme_keyword of an application
- $S_j \square \in S$ - where S is the sum of theme_keywords in DS_S

- w total number theme_keywords defined for application
- R_j percentage value of relevance based on application and TKW similarity found

Algorithm 5: Select datasets using user application and theme_keywords

Procedure:

- for all datasets select a dataset if:
- user application and the theme_keyword is similar to overview quality description of the dataset or
- the theme_keyword is similar to overview quality description of the dataset even if user application is not similar to overview quality description of the dataset.

Input: $U_A, T KW, DS$

- 1: for $i = 1$ to N do
- 2: if $U_A \square \sim OQ$ then
- 3: if $T KW \square \sim OQ$ then
- 4: $DS_S \square \leftarrow DS^i$
- 5: end if
- 6: else
- 7: if $T KW \square \sim OQ$ then
- 8: $DS_S \square \leftarrow DS^i$
- 9: end if
- 10: end if
- 11: $i \leftarrow i+1$
- 12: end for
- 13: return DS_S

The result of algorithm 5 is the set of datasets that the overview quality or description has matching with user application or the corresponding theme_keywords. This datasets used in the process to quantify the application and the theme_keywords matching found in the dataset as shown in algorithm 9:

Algorithm 6: Quantify application name and theme_keyword in DS_S

Procedure: - for all datasets, compare user application with each overview quality description of the dataset. When ever they are similar, set weight of the application as the sum of all the theme_keywords, otherwise set the weight to 0 - for all datasets and for all theme_keyword, if a theme_keyword is similar to overview quality description of the dataset, set the theme_keyword weight to 1, else 0 - return weight assigned for user application and weight assigned for theme keyword

Input: $DS_S, U_A, T KW$

- 1: for DS^i_s to DS^N_s do
- 2: $O^iQ \square \leftarrow \text{get_overview_quality}(DS^iS)$
- 3: if $U_A = O^iQ$ then
- 4: $\Omega_A \square \leftarrow \omega \square / \omega >^M \Omega^T KW_{j=1}$
- 5: else
- 6: $\Omega_A \square \leftarrow 0$
- 7: end if
- 8: for $T KW_j$ to $T KW_M$ do
- 9: if $T KW_j = O^iQ$ then
- 10: $\Omega^T KW \square \leftarrow 1$
- 11: else
- 12: $\Omega^T KW \square \leftarrow 0$
- 13: end if
- 14: end for
- 15: end for
- 16: return $\Omega_A, \Omega^T KW$

The output of algorithm 6 is conversion of subjective matching into quantitative values. As explained before, the weight given for the user application matching should be greater than sum of all the theme_keywords defined for that specific application. This enables to identify datasets that match user application in prior than other datasets selected as candidate datasets based on theme_keywords. The theme_keywords are assigned boolean values as weight to indicate matching is found or not in the dataset where a weight equal to 1 means that matching has been found.

This values again summed up using algorithm 7 as shown below:

Algorithm 7: Compute sum of theme_keywords in dataset

Procedure:

- compute sum of theme_keywords in dataset DS^i theme_keyword

Input: $DS_S, \Omega_T KW$

- 1: for DS^i_S to DS^N_S do
- 2: $S_i \square \leftarrow \sum_{j=1}^M \Omega_j T KW_j$
- 3: end for
- 4: return S

When the process of selecting spatial dataset based on user application, searching datasets by theme_keywords and assigning weight value completed, these values are used to rank spatial datasets that best fit user application. We said that application similarity found in the spatial dataset overview quality have more theme_keyword matching has second priority. Based on this assumption we devise a relevance indicator to inform users how much percent a dataset fits their application. In order to elaborate our approach we assume that there is an application with five theme_keywords and designed algorithm 8 as shown below:

Algorithm 8: Display relevance of DS_S based on application and theme_keywords weight

Procedure:

- For each datasets DS_S filtered by U_A and $T KW$
- Extract the weight Ω_A and S
- If exact matching for user application and the dataset overview quality, then indicates dataset is 100% relevant
- Otherwise calculate the difference between w and s^i to indicate the corresponding relevance

Input: DS_S, w

- 1: for DS^i_S to DS^N_S do
- 2: $\Omega_A \square \leftarrow \text{get_weight_by_}U_A(DS_S)$
- 3: $S \square \leftarrow \text{get_weight_by_}TKW(DS_S)$
- 4: if $\Omega_A = w$ then
- 5: $relevance \square \leftarrow R^i\%$
- 6: else if $w - s^i = 1$ then
- 7: $relevance \square \leftarrow R^i\%$
- 8: else if $w - s^i = 2$ then
- 9: $relevance \square \leftarrow R^i\%$
- 10: else if $w - s^i = 3$ then
- 11: $relevance \square \leftarrow R^i\%$
- 12: else if $w - s^i = 4$ then
- 13: $relevance \square \leftarrow R^i\%$
- 14: else
- 15: $relevance \square \leftarrow R^i\%$
- 16: end if
- 17: end for

Once the final result for application matching is returned from algorithm 8 the relevance indicator can be used to order the datasets as shown in algorithm 9.

Algorithm 9: Rank Selected Data Sets DS_S based on application and theme_keywords

Procedure:

- sort the dataset according to application and sum of theme_keywords

Input: DS_S

- 1: $\Omega_A \square \leftarrow \text{get_app_weight}(DS_S)$
- 2: $S \square \leftarrow \text{get_sum_theme_keyword}(DS_S)$
- 3: $DS_{SS} \square \leftarrow \text{sort } DS_{SS} \text{ desc } \Omega_A, \text{ desc } S$
- 4: return DS_{SS}

FITNESS FOR USE EVALUATION USING QUALITY ELEMENT

To design the fitness for use evaluation of spatial datasets based on quality elements, the spatial datasets need to be extracted based on user extent and application requirement and populated in the system profile. Once the datasets quality description and necessary information are populated in the system, fitness for use evaluation can be performed. In order to compute the fitness of a dataset by comparing the quantitative data quality elements, the measurement unit of each quality element should be adjusted into the same measurement unit.

In this section we address the process of computing the fitness for use evaluation using quantitative data quality elements. to recommend a spatial datasets based on users quality requirements, we design algorithm 13 to compute range for all user quality requirement values. This is because it is not always possible to find spatial dataset that exactly match users requirement. We decided to subtract and add half of user quality requirement value to a user quality requirement value itself for each element to set the minimum and maximum of range.

Algorithm 10: sum of weighted X (dataset relevance to user quality requirement)

Procedure:

- for all datasets weighted boolean value, compute sum of weighted X
- return sum weighted X

Input: $DS, X^w, Q_{DS}, S_j = 0$

- 1: for DS_j to DS_M do
- 2: $X^w \square \leftarrow \text{get_weighted_boolean}(DS)$
- 3: for X^w_{j1} to X^w_{jN} do
- 4: $S_j += X^w_{ji}$
- 5: end for
- 6: end for
- 7: return S

Algorithm 11: Calculate distance of dataset quality from user quality

Procedure:

- for all datasets:
- fetch boolean values of the datasets quality elements
- for each user quality requirements
- for a single dataset, if each boolean value is true, then compute the distance from user quality subelement to dataset quality subelement

Input: DS, Q_R, Q_w

- 1: for DS_j to DS_M do
- 2: $X \square \leftarrow \text{get_boolean}(DS)$
- 3: for Q^{iR} to Q^N do R
- 4: for X_{ji} to X_{jN} do
- 5: /*jth dataset and i to N quality subelements*/
- 6: $X_{ji} = \text{fetch}(X) /*X - \text{boolean value}*/$
- 7: if $X_{ji} = 1$ then
- 8: $d_{ji} \square \leftarrow \square / Q^{iR} \square - Q^i \square \square \square DS$

```

9:     end if
10:  end for
11:  end for
12: end for
13: return  $D$ 

```

Finally using the computed relevance indicator from algorithm 10 and the distance value of each quality element of the dataset from algorithm 11, it is possible to identify the best dataset that fits user quality requirements.

Algorithm 12: Rank datasets DS by relevance based on quality element evaluation

Procedure:

- sort the dataset according to their aggregated weighted boolean X

Input: DS

1: $S \square \leftarrow \text{get_sum_weight_boolean}(DS)$

2: $DSG \square \leftarrow \text{sort } DS \text{ desc } S$

3: return DSG

Variable definition used in Algorithm 16:

- DSG - sorted datasets
- $DSGS$ - sorted DSG by distance
- Q_w - set of weight of quality elements provided by user
- Q^{max} - the maximum weight of quality elements provided by user w
- Q^{name} - the name of the maximum weighted of quality elements w
- D - distance between dataset quality and user quality
- d_{ji} - is distance between dataset quality and user quality in i th column

Algorithm 13: Rank dataset by identifying relevance by distance

Procedure:

- Input datasets sorted by relevance value:

-for all user quality requirement weight

- identify the maximum user quality requirement weight assigned

- identify the name of quality element which have maximum weight

- find the dataset quality element which have the same name

- for each DSG identify the distance value of quality element

- sort the DSG based on the distance value

Input: D, Q_w, DSG

for Q^{iw} to Q^N do w

$Q^{max} \square \leftarrow \text{fetch_max}(Q_w) w$

$Q^{name} \square \leftarrow \text{get_name}(Q^{max})$

if $DSG \sim Q_w^{name}$ then

for DS^j_s to DS^M_s do

$d_{ji} \square \leftarrow \text{fetch}(DSG)$

end for

end if

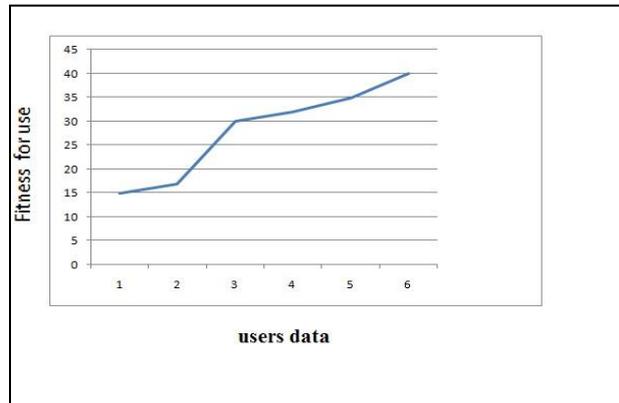
$DSGS \square \leftarrow \text{sort}(DSG) \text{ desc } d_{ji}$

end for

return $DSGS$

II. EXPERIMENTAL RESULTS

The recommendation service result page allows users to access recommended datasets as a response to their requirements. The recommendation process starts first by filtering spatial datasets based on user extent requirement. The datasets that satisfy user spatial extent requirement returned as a candidate datasets. Then application based filtering process starts specifically by comparing the user application to the overview data quality information usage and purpose and continues matching the theme_keywords of the application to the description of datasets



Finally the system recommend spatial datasets with corresponding values that indicates level of relevance

As computed by the fitness for use evaluation logic. The values returned with the recommended datasets enables users to easily observe which dataset best fits which requirements. In order to help users on picking the datasets based on extent, the computed spatial extent will also be visualized on the map.

CONCLUSION

The fundamental approach to determine fitness for use is comparison of users quality requirements and quality of data resources. In order to use fitness for use as a searching criteria in GIS, comprehensive comparison against user quality requirement and detailed quality description of spatial dataset is required. Thus understanding users' view towards spatial data quality and quality description of spatial data resources, gave us an idea on how to design a conceptual model of recommendation system. GIS spatial data quality is a perception or an assessment of data fitness to serve its purpose in a given context and subjective to various applications. Widely accepted expression affirms that spatial data quality is recognized in terms of its specific use and the quality definition given by ISO is accepted in common to describe spatial data quality.

Therefore, we followed the spatial data quality according to ISO standard to represent the spatial data quality and users spatial data search quality requirements in our system. This simplification is required because the standard gives common ground on spatial data quality to evaluate its fitness for use. We also consider OGC catalogue service as a source to extract required datasets quality description. However, data quality description to determine fitness for use should not be limited to the ISO standard. Other factors such as: currency, cost, accessibility of the dataset, dataset granularity, popularity and users opinion about the dataset need to be included.

REFERENCES

- [1] E. and B. VaBeur. How to select the best dataset for a task. In *Proceedings of 3rd International Symposium on Spatial Data Quality (ISSDQ'04)*, pages 197-206, 2004.
- [2] ISO/TC 211. Text of 19113 geographic information - quality principles, as sent to the iso central secretariat for registration as fdis, 2002.
- [3] A. Agumya and G.J. Hunter. A risk-based approach to assessing the fitness for use of spatial data. *URISA Journal*, 11(1):33-44, 1999.
- [4] Victorian Spatial Council. Spatial Information Data Quality Guidelines. Victorian Spatial Council, 2009.
- [5] M.A. Gebesilassie, I. Ivánová, and J. Morales. User profiles for data quality models. Master's thesis, University of Twente Faculty of Geo-Information and Earth Observation ITC, 2011.
- [6] ISOTC211. Revised text of 19115 Geographic information - Metadata, as sent to the ISO Central Secretariat for registration as FDIS, 2003.
- [7] ISOTC211. Text of 19114 geographic information - quality evaluation procedures as sent to the iso central secretariat for publication, 2003.
- [8] A.U. Frank, E. Grum, and B. VaBeur. Procedure to select the best dataset for a task. *Geo- graphic Information Science*, pages 81-93, 2004.

- [9] K.T. Huang, Y.W. Lee, and R.Y. Wang. *Quality information and knowledge*, volume 141. Prentice Hall PTR, 1999.
- [10] A. Zargar and R. Devillers. An operation-based communication of spatial data quality. In *2009 International Conference on Advanced Geographic Information Systems & Web Services*, pages 140-145. IEEE, 2009.

Second Derivative Free Modification with a Parameter For Chebyshev's Method

J. Jayakumar¹ and P. Jayasilan

Department of Mathematics, Pondicherry Engineering College, Pondicherry 605 014, India.

ABSTRACT:

Chebyshev's method has not got much attention in recent years compared to Newton's method. Some of the papers which appeared recently discuss modification of Chebyshev's method free from second order derivative. In this paper, we have modified Chebyshev's method by suitably approximating second order derivative using Taylor's Series. The proposed method requires evaluation of only three functions and still maintains cubic convergence. For a particular choice of a parameter in the method, we get fourth order convergence. Examples are provided to show the efficiency of the method with classical Chebyshev's method and few other cubic convergent methods.

KEYWORDS: Chebyshev's Method, Iterative Method, Non-linear equation, Second derivative free method, Cubic convergence.

I. INTRODUCTION

One of the most important and challenging problem in applied mathematics and engineering is to find an approximate solution of the nonlinear equation

$$f(x) = 0 \quad (1)$$

where $f: D \subset \mathbb{R} \rightarrow \mathbb{R}$ for an open interval D is a scalar function. Newton's method is one of the famous iterative methods to solve (1). It is well known that it has quadratic convergence. The iterative formula of Newton's method is given by

$$y_n = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2 \dots \quad (2)$$

By improving Newton's method, Chebyshev [8] derived a cubic convergent iterative method which requires computing of second derivative of the function (1). Without evaluating second derivative of $f(x)$, cubic convergence was first established by the Arithmetic mean Newton's method [1, 16]. In [1, 4, 11, 14] third order accuracy is proved by using midpoint Newton's method. Modifications in the Newton's method using harmonic mean were suggested in [1, 7, 11, 14]. Nedzhibov [13] gave several classes of iterative methods using different quadrature rules for solving nonlinear equations. Different variants based on Simpson formula on Newton's theorem were proposed in [1, 2, 5, 18]. Zhou [17] gave a class of Newton's methods based on power means on the trapezoid formula. Tibor et al [12] gave third order accurate Newton's modification by using geometric mean. Jisheng et al [10] proposed a uniparametric Chebyshev type method free from second derivative. Hecceg et al [6] presented a method for constructing new third order methods for solving (1) in which Halley's and Chebyshev's methods are special cases. Esmaeili et al [3] gave a uniparametric modification of Chebyshev's method. Different variants of Chebyshev's method with optimal order of convergence were proved in [15]. Jayakumar et al [9] modified Newton's method using the harmonic mean instead of arithmetic mean on the Simpson's formula. The methods presented in all the above papers require only the first derivative of $f(x)$ and establishes third order convergence.

In this paper, we propose a modification in the Chebyshev's method [8] similar to the one given in [10] which is free from second order derivative of $f(x)$ by considering a suitable approximation of $f''(x)$ using Taylor's Series. The proposed new method has the advantage of evaluating only the first derivative of $f(x)$, less number of iterations and third order accuracy. With an additional condition on the parameter k , fourth order accuracy is obtained. In Section 2, we present some definitions related to our study. In Section 3, some known third order variants of Newton's method are discussed. Section 4 presents the new method and its analysis of convergence. Finally, Section 5 gives numerical results and discussion.

II. DEFINITIONS

Definition 2.1 [16]: Let $\alpha \in R, x_n \in R, n = 0, 1, 2, \dots$. Then the sequence $\{x_n\}$ is said to converge to α if $\lim_{n \rightarrow \infty} |x_n - \alpha| = 0$. If, in addition, there exist a constant $c \geq 0$, an integer $n_0 \geq 0$ and $p \geq 0$ such that for all $n > 0$, $|x_{n+1} - \alpha| \leq c |x_n - \alpha|^p$, then $\{x_n\}$ is said to converge to α with order at least p . If $p = 2$ or 3 , the convergence is said to be quadratic or cubic.

Definition 2.2 [8]: The efficiency index of an iterative method in the sense of Traub is defined by the equation $E^* = p^{1/n}$, where p is the order of the method and n is the total number of function and derivative evaluations at each step of the iteration.

Definition 2.3 [16]: Let α be a root of the function (1) and suppose that x_{n-1}, x_n and x_{n+1} are three successive iterations closer to the root α . Then, the computational order of convergence (COC) ρ can be approximated using the formula $\rho \approx \frac{\ln|(x_{n+1}-\alpha)/(x_n-\alpha)|}{\ln|(x_n-\alpha)/(x_{n-1}-\alpha)|}$.

III. KNOWN THIRD ORDER METHODS

Let α be a simple root of a sufficiently differentiable function $f(x)$. Consider the numerical solution of $f(x) = 0$. From Newton's theorem, we have

$$f(x) = f(x_n) + \int_{x_n}^x f'(t) dt. \tag{3}$$

Arithmetic Mean Newton's method (AN): By using the trapezoidal rule in (3), we obtain

$$\int_{x_n}^x f'(t) dt \approx \frac{(x-x_n)}{2m} \left[f'(x_n) + 2 \sum_{i=1}^{m-1} f' \left(x_n - \frac{if(x_n)}{mf'(x_n)} \right) + f'(x) \right] \tag{4}$$

From equations (3) and (4), a new approximation x_{n+1} for x is obtained and for $m = 1$, we get the AN [1, 16]

$$x_{n+1} = x_n - \frac{2f(x_n)}{f(x_n) + f'(y_n)}, n = 0, 1, 2 \dots \tag{5}$$

Since equation (5) is implicit, we can overcome this implicit nature by the use of Newton's iterative step (2).

Harmonic Mean Newton's method (HN): If we use the harmonic mean instead of arithmetic mean in (5), we obtain harmonic mean Newton's method [1, 7, 11, 14]:

$$x_{n+1} = x_n - \frac{f(x_n)(f'(x_n) + f'(y_n))}{2f'(x_n)f'(y_n)}, n = 0, 1, 2 \dots \tag{6}$$

where y_n is calculated from (2).

Midpoint Newton's method (MN): If the integral in (3) is approximated using the midpoint integration rule instead of trapezoidal rule, we get the midpoint Newton's method [1, 4, 11, 14]:

$$x_{n+1} = x_n - \frac{f(x_n)}{f' \left(\frac{x_n + y_n}{2} \right)}, n = 0, 1, 2, \dots \tag{7}$$

where y_n is calculated from (2).

IV. MODIFICATION OF CHEBYSHEV'S METHOD AND ITS ANALYSIS OF CONVERGENCE

The classical Chebyshev's method (CCM) is given by

$$x_{n+1} = x_n - \left(1 + \frac{1}{2} \left[\frac{f(x_n) f''(x_n)}{f'(x_n)^2} \right] \right) \frac{f(x_n)}{f'(x_n)}. \tag{8}$$

Use Taylor's series to approximate $f''(x_n)$ in (8) with $\frac{f'(y_n) - f'(x_n)}{y_n - x_n}$, where y_n is given by

$$y_n = x_n - k \frac{f(x_n)}{f'(x_n)}, k \neq 0, n = 0, 1, 2, \dots \tag{9}$$

Thus equation (8) is converted into

$$x_{n+1} = x_n - \left(1 + \frac{1}{2} \left[\frac{f(x_n)}{f'(x_n)} \right] \left[\frac{1}{f'(x_n)} \right] \left[\frac{f'(y_n) - f'(x_n)}{y_n - x_n} \right] \right) \frac{f(x_n)}{f'(x_n)}. \quad (10)$$

We can modify equation (10) by using (9) as below

$$x_{n+1} = x_n - \left(1 - \frac{1}{2} \left[\frac{f(x_n)}{f'(x_n)} \right] \left[\frac{1}{f'(x_n)} \right] \left[\frac{f'(y_n) - f'(x_n)}{-k \frac{f(x_n)}{f'(x_n)}} \right] \right) \frac{f(x_n)}{f'(x_n)},$$

where the parameter $k \neq 0$. Finally, we simplify the above equation to obtain the new method

$$x_{n+1} = x_n + \left(-1 + \frac{1}{2k} \left[\frac{f'(y_n) - f'(x_n)}{f'(x_n)} \right] \right) \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2, \dots \quad (11)$$

Theorem 4.1 Assume that the function $f: D \subset \mathbb{R} \rightarrow \mathbb{R}$ has a simple root $\alpha \in D$, where D is an open interval. If $f(x)$ is sufficiently differentiable in the interval D , then the methods defined by (11) converge cubically to α in a neighbourhood of α , for all values of the parameter $k \neq 0$. Moreover, if $f''(\alpha) = 0$ and $k = 2/3$ then the method has fourth order convergence.

Proof. Let α be a simple zero of function $f(x) = 0$. (That is, $f(\alpha) = 0$ and $f'(\alpha) \neq 0$).

Let $e_n = x_n - \alpha$ and $d_n = y_n - x_n$, where y_n is defined by (9). By expanding $f(\alpha)$ by Taylor's series about x_n , we obtain

$$f(\alpha) = f(x_n) - e_n f'(x_n) + \frac{1}{2} f''(x_n) e_n^2 - \frac{1}{6} f'''(x_n) e_n^3 + O(e_n^4). \quad (12)$$

Taking into account $f(\alpha) = 0$, we have from (12)

$$f(x_n) = e_n f'(x_n) - \frac{1}{2} f''(x_n) e_n^2 + \frac{1}{6} f'''(x_n) e_n^3 + O(e_n^4). \quad (13)$$

Let us denote $c_j = \frac{f^{(j)}(x_n)}{n! f'(x_n)}$, $j = 2, 3, \dots$. Dividing equation (13) by $f'(x_n)$, we get

$$\frac{f(x_n)}{f'(x_n)} = e_n - C_2 e_n^2 + C_3 e_n^3 - C_4 e_n^4 + O(e_n^5). \quad (14)$$

From (9) and (14), we have

$$d_n = y_n - x_n = -k \frac{f(x_n)}{f'(x_n)} = -k e_n + k C_2 e_n^2 - k C_3 e_n^3 + k C_4 e_n^4 + O(e_n^5). \quad (15)$$

Using Taylor's expansion for $f'(y_n)$ about x_n , we get

$$f'(y_n) = f'(x_n) + f''(x_n) d_n + \frac{f^{(3)}(x_n)}{2} d_n^2 + \frac{f^{(4)}(x_n)}{6} d_n^3 + \frac{f^{(5)}(x_n)}{24} d_n^4 + O(d_n^5). \quad (16)$$

From (15) and (16), we get

$$\begin{aligned} f'(y_n) - f'(x_n) &= f''(x_n) k [-e_n + C_2 e_n^2 - C_3 e_n^3 + C_4 e_n^4] + \frac{f^{(3)}(x_n)}{2} k^2 [e_n^2 - 2C_2 e_n^3 + (C_2^2 + 2C_3) e_n^4 \\ &\quad - 2(C_4 + C_2 C_3) e_n^5] + \frac{f^{(4)}(x_n)}{6} k^3 [-e_n^3 + 3(C_2 e_n^4 - C_2^2 e_n^5 + C_3 e_n^5)] + \frac{f^{(5)}(x_n)}{24} k^4 [e_n^4 - 4C_2 e_n^5] + O(e_n^6). \end{aligned} \quad (17)$$

Dividing (17) by $f'(x_n)$ and the result is multiplied by $\frac{1}{2k}$ to get

$$\frac{1}{2k} \left(\frac{f'(y_n) - f'(x_n)}{f'(x_n)} \right) = -e_n C_2 + e_n^2 \left(C_2^2 + \frac{3}{2} k C_3 \right) - e_n^3 [C_2 C_3 (1 + 3k) + 2k^2 C_4] + e_n^4 \left[3k \left(\frac{1}{2} C_2^2 C_3 + 2C_3^2 \right) + C_2 C_4 (1 + 6k^2) + \frac{5}{2} k^3 C_5 \right] + O(e_n^5)$$

(18)

From (14) and (18), we have

$$\left[\frac{1}{2k} \left(\frac{f'(y_n) - f'(x_n)}{f'(x_n)} \right) \right] \frac{f(x_n)}{f'(x_n)} = -e_n^2 C_2 + e_n^3 (2C_2^2 + \frac{3k}{2} C_3) - e_n^4 (C_2 C_3 (2 + \frac{9}{2} k) + C_2^3 + 2k^2 C_4) + O(e_n^5).$$

Again from (14) and the above equation, we get

$$-\frac{f(x_n)}{f'(x_n)} + \frac{1}{2k} \left[\left(\frac{f'(y_n) - f'(x_n)}{f'(x_n)} \right) \right] \frac{f(x_n)}{f'(x_n)} = -e_n + e_n^3 \left(2C_2^2 + \left(\frac{3k}{2} - 1 \right) C_3 \right) - e_n^4 \left(C_2 C_3 (2 + \frac{9}{2} k) + C_2^3 + C_4 (2k^2 - 1) \right) + O(e_n^5)$$

(19)

Using the relation $e_n = x_n - \alpha$ in (11), we have

$$e_{n+1} = e_n - \frac{f(x_n)}{f'(x_n)} + \frac{1}{2k} \left[\left(\frac{f'(y_n) - f'(x_n)}{f'(x_n)} \right) \right] \frac{f(x_n)}{f'(x_n)}$$

(20)

Finally, using (19) on (20), we obtain

$$e_{n+1} = e_n^3 \left[2C_2^2 + C_3 \left(\frac{3k}{2} - 1 \right) \right] - e_n^4 \left[C_2^3 + C_2 C_3 \left(2 + \frac{9k}{2} \right) + C_4 (2k^2 - 1) \right] + O(e_n^5)$$

(21)

From equation (21), we conclude that the methods defined by (11) are cubically convergent for all the values of $k \neq 0$. Also, taking $f''(x_n) = 0$ (i.e. $C_2 = 0$) and $k = 2/3$ in equation (21), we get fourth order convergence.

V. NUMERICAL EXAMPLES AND DISCUSSION

In this section, we present the results of numerical calculations to compare the efficiency of the present method (MCM). We compare MCM with Newton's method (NM), classical Chebyshev method (CCM) and other cubic convergent Newton-type methods given in section III. Number of iterations and Computational Order of Convergence (COC) for different functions is given in **Table I**. Numerical computations are worked out in the MATLAB software with double precision accuracy and the results are presented below.

Stopping criterion for the iterations has been taken as $|x_{n+1} - x_n| + |f(x_{n+1})| < \epsilon$, where $\epsilon = 10^{-14}$. The following table gives functions that are taken as examples:

$f(x)$	Root (α)	$f(x)$	Root (α)
$f_1(x) = \arctan(x)$	0	$f_4(x) = \ln(x^3 + x + 1)$	0
$f_2(x) = e^{-x} \sin x + \log(1 + x^2)$	0	$f_5(x) = x^2 \sin(x) - \cos(x)$	6.3083089552381
$f_3(x) = x^3 - 9x^2 + 28x - 30$	3	$f_6(x) = (x + 2)e^{-x} + x$	-1.6878939988

Discussion: In this work, a single parameter family of Chebyshev-type method free from second order derivatives for solving non-linear equation is presented. Third-order convergence is established for the proposed method for all values of the parameter $k \neq 0$. Also, we get fourth order convergence when $k = 2/3$ and $C_2 = 0$ which is displayed for the examples $f_1(x)$, $f_2(x)$ and $f_3(x)$. The efficiency index for MCM is $E^* = 1.4422$ and for the value of $k = 2/3$ and $C_2 = 0$, $E^* = 1.5874$ which are very good compared to Newton's method which has the efficiency $E^* = 1.4142$. This shows that MCM is preferable to Newton's method, especially when the computational cost of the first derivative is not more than that of the function itself. These proposed methods are also superior as there is no evaluation of second derivative compared to its classical predecessor, Chebyshev's method. From numerical examples, we have shown that these new methods have very good practical utility.

Table I

$f(x)$	x_0	Number of iterations							COC						
		NM	AN	MN	CC M	MCM k=1	MCM k=1/2	MCM k=2/3	NM	AN	MN	CCM	MCM k=1	MCM k=1/2	MCM k=2/3
$f_1(x)$	1.3	8	6	5	8	6	NC	NC	3.0	2.9	3.0	3.0	2.9	NC	NC
	1	6	5	5	6	5	6	5	2.9	2.9	3.0	3.0	2.9	3.3	4.9
	0.5	5	4	4	5	4	4	4	2.9	2.9	3.1	3.0	2.8	3.5	4.9
	-1	6	5	5	6	5	6	5	2.9	2.9	3	3.0	2.9	3.3	4.9
$f_2(x)$	1.3	5	4	5	5	5	5	4	3.0	2.2	2.9	2.56	2.9	2.8	4.0
	1	5	4	4	5	4	4	4	3.0	2.6	2.8	2.79	2.5	3.0	3.6
	0.5	4	4	4	4	4	4	4	2.4	2.6	2.7	2.78	2.6	2.7	3.8
	-1	6	5	5	5	5	5	5	3.0	3.1	3.0	3.07	3.1	2.8	4.4
$f_3(x)$	2	6	5	5	5	5	5	5	2.9	2.9	2.7	2.52	3.1	3.0	4.8
	2.5	7	6	5	6	6	6	7	2.9	2.9	2.9	2.97	3.0	2.8	4.8
	3.5	8	6	6	7	7	7	6	3.0	2.9	2.9	1.51	2.6	1.5	4.8
	1	9	7	7	7	7	7	7	2.9	2.9	2.9	2.78	3.0	2.9	4.8
$f_4(x)$	1.3	6	4	5	5	4	5	4	2.0	1.8	3.1	2.9	1.9	2.9	2.2
	1	6	4	4	5	5	4	4	2.0	2.0	2.6	3.0	2.9	3.5	2.3
	0.5	6	4	4	4	4	4	4	2.0	2.5	3.1	2.5	2.4	2.7	2.4
	-1	8	7	6	7	7	7	7	2.0	3.0	3.0	2.98	3.0	3.0	2.9
$f_5(x)$	4	6	4	4	5	4	5	4	1.9	2.4	2.5	2.9	2.2	2.9	2.3
	1	5	4	4	4	4	4	4	1.9	2.9	2.9	2.9	2.9	2.9	2.9
	0.5	7	5	5	7	9	8	7	2.0	3.0	3.0	3.0	3.1	2.9	2.8
	-4	6	4	5	4	5	5	5	1.9	2.1	2.9	2.05	2.9	2.9	2.9
$f_6(x)$	-3	8	6	6	8	6	6	6	1.9	2.8	2.9	1.99	2.9	2.9	2.9
	-2	6	4	4	6	5	5	5	1.9	2.8	2.9	1.99	2.9	2.9	2.9
	-1	9	7	6	13	15	13	D	2.0	3.0	3.0	1.99	2.9	2.9	D

NC – Not Convergent, D – Divergence, x_0 – Initial point and COC – Computational Order of Convergence

REFERENCES

- [1] D.K.R. Babajee, M.Z. Dauhoo, An analysis of the properties of the variants of Newton’s method with third order convergence, Applied Mathematics and Computation 183 (2006) 659-684.
- [2] A. Cordero, J. R. Torregrosa, Variants of Newton’s Method using fifth-order quadrature formulas, Applied Mathematics and Computation 190 (2007) 686-698.
- [3] H. Esmaili and A.N. Rezaei, A uniparametric family of modifications for Chebyshev’s method, Lecturas Matematicas 33 (2) (2012) 95-106.
- [4] M. Frontini, E. Sormoni, Some variants of Newton’s method with third order convergence, Applied Mathematics and Computation 140 (2003) 419-426.
- [5] V.I. Hasanov, I.G. Ivanov, G. Nedjibov, A New Modification of Newton’s Method, Applications of Mathematics in Engineering 27 (2002) 278 -286.
- [6] Djordje Herceg, Dragoslav Herceg, A method for obtaining third-order iterative formulas, Novi Sad Journal of Mathematics 38 (2) (2008) 195-207.
- [7] H.H.H. Homeier, On Newton type methods with cubic convergence, Journal of Computational and Applied Mathematics 176 (2005) 425-432.
- [8] M.K. Jain, S.R.K. Iyengar, R.K. Jain, Numerical Methods for Scientific and Engineering Computation, New Age International, 6th edition, 2012.
- [9] J. Jayakumar, Kalyanasundaram M., Modified Newton’s method using harmonic mean for solving nonlinear equations, IOSR Journal of Mathematics 7 (4) (2013) 93-97.
- [10] J. Kou, Y. Li, X.Wang, A uniparametric Chebyshev-type method free from second derivatives, Applied Mathematics and Computation 179 (2006) 296 – 300.
- [11] J. Kou, Y. Li, X.Wang, Third-order modification of Newton’s method, Journal of Computational and Applied Mathematics 205 (2007) 1 – 5.
- [12] T. Lukic, N.M. Ralevic, Geometric mean Newton’s method for simple and multiple roots, Applied Mathematics Letters 21 (2008) 30-36.
- [13] G. Nedzhibov, On a few iterative methods for solving nonlinear equations. Application of Mathematics in Engineering and Economics’28, in: Proceeding of the XXVIII Summer school Sozopol’ 02, pp.1-8, Heron press, Sofia, 2002.
- [14] A.Y. Ozban, Some New Variants of Newton’s Method, Applied Mathematics Letters 17 (2004) 677-682.
- [15] Ramandeep Behl, V. Kanwar, Variants of Chebyshev’s method with optimal order of convergence, Tamsui Oxford Journal of Information and Mathematical sciences 29(1) (2013) 39-53.
- [16] S. Weerakoon. T.G.I. Fernando, A Variant of Newton’s method with accelerated third-order convergence, Applied Mathematics Letters 13 (8), (2000) 87-93.
- [17] Zhou Xiaojian, A class of Newton’s methods with third-order convergence, Applied Mathematics Letters 20 (2007)1026 – 1030.
- [18] F. Zafar, N.A. Mir, A generalized family of quadrature based iterative methods, General Mathematics 18(4) (2010) 43-51.

Protecting Frequent Item sets Disclosure in Data Sets and Preserving Item Sets Mining

¹, Shaik Mahammad Rafi,², M.Suman ,M.Tech³, Majjari Sudhakar,⁴, M.Tech

⁴, P.Ramesh⁵, P.Venkata Ramanaih.,M.Tech

¹PG (M.Tech) Student in CSE Department, Global College of Engineering and Technology, Kadapa, YSR (D.t).

^{2,3,4}Assistant Professors in CSE Department, MRRITS, Udayagiri, SPSR Nellore(D.t).

⁵ Assistant Professor in CSE Department, GCET, Kadapa, YSR(D.t).

ABSTRACT:

Based on the network and data mining techniques, the protection of the confidentiality of sensitive information in a database becomes a critical issue to be resolved. Association analysis is a powerful and popular tool for discovering relationships hidden in large data sets. The relationships can be represented in a form of frequent itemsets or association rules. One rule is categorized as sensitive if its disclosure risk is above some given threshold. Privacy-preserving data mining is an important issue which can be applied to various domains, such as Web commerce, crime reconnoitering, health care, and customer's consumption analysis.

The main approach to hide sensitive frequent itemsets is to reduce the support of each given sensitive itemsets. This is done by modifying transactions or items in the database. However, the modifications will generate side effects, i.e., nonsensitive frequent itemsets falsely hidden (the loss itemsets) and spurious frequent itemsets falsely generated (the new itemsets). There is a trade-off between sensitive frequent itemsets hidden and side effects generated. Furthermore, it should always take huge computing time to solve the problem.

In this study, we propose a novel algorithm, FHSFI, for fast hiding sensitive frequent itemsets (SFI). The FHSFI has achieved the following goals: 1) all SFI can be completely hidden while without generating all frequent itemsets; 2) limited side effects are generated; 3) any minimum support thresholds are allowed, and 4) only one database scan is required.

I. INTRODUCTION

The data mining technologies have been an important technology for discovering previously unknown and potentially useful information from large data sets or databases. They can be applied to various domains, such as Web commerce, crime reconnoitering, health care, and customer's consumption analysis. Although these are useful technologies, there is also a threat to data privacy. For example, the association rule analysis is a powerful and popular tool for discovering relationships hidden in large data sets. Therefore, some private information could be easily discovered by this kind of tools. The protection of the confidentiality of sensitive information in a database becomes a critical issue to be resolved.

The relationships discovered from a database can be represented in a form of frequent itemsets or association rules. One rule is categorized as sensitive if its disclosure risk is above some given threshold. With an association analyzer, if an itemset with support above a given minimal support, we call the itemset as a frequent itemset.

The problem for finding an optimal sanitization of a source database with association rule analysis has been proven to be NP-Hard [1]. In [2,3,4,5] the authors presented different heuristic algorithms that modify transactions via inserting or deleting items for hiding sensitive rules or itemsets.

Vassilios S. Verykios et al. [2] presented algorithms to hide sensitive association rules, but they generate high side effects and require multiple database scans. Instead of hiding sensitive association rules, Shyue-Liang Wang [3] proposed algorithms to hide sensitive items. The algorithm needs less number of database scans but the side effects generated is higher. Ali Amiri [4] also presented heuristic algorithms to hide sensitive items. Finally, Yi-Hung Wu et al. [5] proposed a heuristic method that could hide sensitive association rules with limited side effects. However, it spent a lot of time on comparing and checking if the sensitive rules are hidden and if side effects are produced. Besides, it could fail to hide some sensitive rules in some cases.

In this study, we propose a novel algorithm, FHSFI for fast hiding sensitive frequent itemsets (SFI). The FHSFI has achieved the following goals: 1) all SFI can be completely hidden while without generating all frequent itemsets; 2) limited side effects are generated; 3) any minimum support thresholds are allowed, and 4) only one database scan is required.

The remainder of this paper is organized as follows: Section 2 presents the problem formulation and notations. In Section 3, we introduce the concept of the proposed algorithm for fast hiding sensitive frequent itemsets and giving examples to illustrate the proposed algorithm. Section 4 is the experimental results which present the performance and various side effects of the proposed algorithm. Section 5 is the conclusion and further work.

II. PROBLEM FORMULATION AND NOTATIONS

In Table 1, we summarize the notations used hereafter in this paper. Let I be a set of items in a transaction database D .

And let $I = \{i_1, i_2, \dots, i_m\}$; $D = \{t_1, t_2, \dots, t_n\}$, where every transaction t_i is a subset of I , i.e. $t_i \subseteq I$. An example database is shown in Table 2. Let X be a set of items in I . If $X \subseteq t_i$, we say that the transaction t_i supports X . There are nine items, $|I|=9$, be minimized.

Table 1. Definitions of variables used in this paper

Variable	Definition
D	the original database
D'	the released database which is transformed from D
U	the sets of frequent item sets generated from D
U'	the sets of frequent item sets generated from D'
t_i	a transaction in Database D
$ t_i $	the number of items in t_i
TID	a unique identifier of each transaction
SFI	the set of sensitive frequent itemsets to be hidden
$SFI.t_j$	a sensitive frequent itemset in the SFI
$\ \cdot \ $	the support count of an itemset, i.e., the number of transactions that support the itemset
w_i	prior weight of t_i
PWT	a table for storing TID and w_i for each transaction in an order decreasing by w_i
MIC_i	the maximal number of itemsets in SFI that contain an item i_k , where $i_k \in t_i, SFI.t_j \subseteq t_i$
$SFI.t_i$	transaction to be modified

and five transactions, $|D|=5$, in the database. The support of itemset X can be computed by equation (1). An association rule is an implication of the form $X \rightarrow Y$, where $X \subset I, Y \subset I$ and $X \cap Y = \emptyset$. A rule $X \rightarrow Y$ will be extracted from a database if

- 1) $support(X \cup Y) \geq \min_support$ (a given minimum support threshold) and
- 2) $confidence(X \cup Y) \geq \min_confidence$ (a given minimum confidence threshold),

where $support(X \cup Y)$ and $confidence(X \cup Y)$ are given by equations (2) and (3), .

$$support(X) = \|X\| / |D| \tag{1}$$

$$support(X \cup Y) = \|X \cup Y\| / |D| \tag{2}$$

$$confidence(X \cup Y) = \|X \cup Y\| / |X| \tag{3}$$

Table 2.
Database D

TID	Transaction
1	1,2,4,5,7

2	1,4,5,7
3	1,4,6,7,8
4	1,2,5,9
5	6,7,8

Table 3.Frequent Itemsets

Itemset	Support
1	80%
4	60%
5	60%
7	80%
1,4	60%
1,5	60%
1,7	60%
1,4,7	60%
4,7	60%

In equation (1), $\|X\|$ denotes the number of transactions in the database that contains the itemset X , and $|D|$ denotes the number of the transactions in the database D . If $\text{support}(X) \geq \text{min_support}$, we call X as a *frequent itemset*. Table 3 shows the frequent itemsets for a given $\text{min_support} = 60\%$.

For the example $X = \{1,4,7\}$, since $X \subseteq t_1, X \subseteq t_2$ and $X \subseteq t_3$, we obtain $\|X\|=3$. Therefore, $\text{support}(1,4,7)=60\%$. Using the form $X \rightarrow Y$ (support, confidence) for association rules, the rules generated from the above itemset $\{1,4,7\}$ can be described as $1 \rightarrow 4,7$ (60%,75%), $4 \rightarrow 1,7$ (60%,100%), $7 \rightarrow 1,4$ (60%,75%), $1,4 \rightarrow 7$ (60%,100%), $1,7 \rightarrow 4$ (60%,100%) and $4,7 \rightarrow 1$ (60%,100%).

Figure 1 shows the relationships among the sets, U, U' , and SFI. The study goal is to hide all SFI and to minimize the loss itemsets. That is, $U' \cap \text{SFI} = \emptyset$ and the set $U - U' - \text{SFI}$ should be minimized.

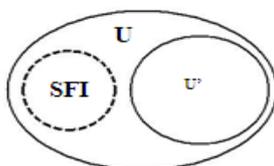


Figure 1. The relationships among the sets, U, U' , and SFI

III. THE PROPOSED ALGORITHM

We now demonstrate the algorithm, FHSFI. Given D, SFI , and min_support , the algorithm is to generate a database to be released, D' , in which the sensitive frequent itemsets are hidden and the side effects generated are minimized.

The sketch of the FHSFI algorithm is shown in Figure 2, which can be depicted as the following stages.

```

FHSFI();
Input:  $D, \text{SFI}, \text{min\_support};$ 
Output:  $D';$ 

Stage 1
1 For each transaction  $t_i$  in the Database  $D$  Do
2   Begin
3     If exist any SFI  $t_i$  (sensitive frequent itemsets) supported by  $t_i$  then
4       Begin
5          $\|\text{SFI}_{t_i}\| := \|\text{SFI}_{t_i}\| + 1;$ 
6          $\text{MIC}_i :=$  the maximum item count from Heuristic( $i, \text{SFI}$ );
7         Compute the prior weight  $w_i$  for each  $t_i$  by the function;
8          $w_i = 1 / [2^{(\|t_i\| - 1)} / \text{MIC}_i];$ 
9         Store the TID and the  $w_i$  in PWT;
10      End;
11 End;

Stage 2
11 While SFI is not empty ( $\neq \emptyset$ ) do
12   Begin
13     Select a TID from PWT with maximal  $w_i$ ;
14     Determine which item in TID will be modified according to
15     Heuristic(TID, SFI);
16     Modify the item;
17     Modify  $w_{\text{TID}}$  of the TID and insert the TID into the PWT;
18   For each SFI  $t_j$  that contains the modified item Do
19     Begin
20        $\|\text{SFI}_{t_j}\| := \|\text{SFI}_{t_j}\| - 1;$ 

```

Stage 2 repeats to modify transitions one-by-one until all SFI have been hidden. The order of the transaction modifications is according to the prior weight associated with a transition. The following tasks are repeated until SFI is empty.

- Select a transaction t_k from PWT such that w_k is maximal.
- Select the item to be deleted, according to the heuristic shown in Figure 4, and delete it.
- Recompute w_k after modifying each item, and then insert it into the PWT in the maintained order.
- Subtract 1 from $\|SFI.t_j\|$ if $SFI.t_j$ contains the deleted item and is supported by t_k .
- Remove $SFI.t_j$ from SFI, if the $(\|SFI.t_j\| / |D|) < \text{min_support}$.

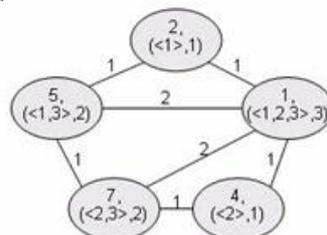


Figure 3. The correlation between t_1 and SFI

```

IF  $\|SFI.t_j\| / |D| < \text{min\_support}$ 
20 then
21   Remove  $SFI.t_j$  from SFI;
22 End;
23 End;
    
```

Figure 2. The pseudo code of the FHSFI algorithm

In stage 1, FHSFI scans database once while collects all useful information about the correlation with SFI for each

Table 4.

An example of sensitive frequent itemsets, SFI

	Itemset
1	1,2,5
2	1,4,7
3	1,5,7
4	6,8

Table 5.

The support count for each itemset in SFI

	Itemset	$\ \cdot\ $
1	1,2,5	2
2	1,4,7	3
3	1,5,7	2
4	6,8	2

transaction, including $\|SFI.t_j\|$ and w_i . The $\|SFI.t_j\|$ is used for checking if $SFI.t_j$ has been hidden. The w_i is a prior weight of a transaction t_i , which provides a heuristic for estimating side effects and can be computed by equation (4).

$$w_i = 1 / [2^{(|t_i|-1)} / MIC_i].$$

Table 4 shows an example of sensitive frequent itemset. Let $t_1 = \{1,2,4,5,7\}$, which supports $SFI.t_1$, $SFI.t_2$ and $SFI.t_3$. As shown in Figure 3 the correlation between t_1 and the SFI can be represented by a graph $G=<V,E>$. Each node is for an item i_k in t_1 ; the weight associated with each edge in E denotes the number of the itemsets in SFI that contain the both adjacent nodes connected by the edge. Each node can be represented as $(\{SFI.t_j \mid SFI.t_j \subseteq t_i, i_k \in SFI.t_j\}, \text{item_count}_{SFI.t_i})$. For example, the node $\langle \{1,2,3\}, 3 \rangle$ for item '1' indicates that three itemsets in SFI that contain the item '1', namely the $SFI.t_1$, $SFI.t_2$, and $SFI.t_3$. As shown in Figure 3, item '1' has the maximum $\text{item_count}_{SFI.t_i}$ which is equal to 3. Hence, we obtain $MIC_1 = 3$ and $w_1 = 3/16$.

Figure 4 shows the heuristic procedure for determining which item to be modified and for computing MIC for transaction t_i .

```

Heuristic ();
Input: TID, SFI;
Output: the item to be modified, MICi;
1 Begin
2 For each SFI.t in SFI do
3 Begin
4 If the transaction tTID fully supports SFI.t; then
5 Begin
6 For each item SFI.tj.i in SFI.tj Do
7 item_countSFI.tj.i = item_countSFI.tj.i + 1;
8 End;
9 End;
10 Select the SFI.t.i with maximum item_count as the item of tTID to be midified;
11 Return(SFI.tj.i, item_count);
12 End;
    
```

Fig:pseudo code of heuristic procedure

Table 6.

The MIC and prior weight for each transaction in D

TID	Transaction	t _i	MIC	w
1	1,2,4,5,7	5	3	3/16
2	1,4,5,7	4	2	2/8
3	1,4,6,7,8	5	1	1/16
4	1,2,5,9	4	1	1/8
5	6,7,8	3	1	1/4

Table 7.The example PWT

	TID	w
1	2	2/8
2	5	1/4
3	1	3/16
4	4	1/8
5	3	1/16

Table 8. Experiment results for |SFI|=5

D	CPU time(ms)	U	U'	#loss itemsets	#modified entries
5000	326.6	439	428.6	5.4	143
10000	454.2	417	406.4	5.6	307.2
15000	701	426	415.6	5.4	513
20000	905	442	431	6	711.6
25000	1183.6	432	421.2	5.8	902.8
30000	1502	443	432.4	5.6	863.8

Now, we use the following example for illustrating the proposed algorithm FHSFI.

Example 1. Given D, SFI, as shown in Tables 2 and 4, and min_support = 40%. As shown in Table 5, the support count for each SFI.t can be obtained from D and SFI. For example, SFI.t₂, {1,4,7}, is supported by t₁, t₂, and t₃, so ||SFI.t₂|| = 3. Table 6 lists the length, MIC, and the prior weight for each transaction in the database. The PWT, as shown in Table 7, can be obtained by sorting Table 6 in the decreasing order by w. Then, the first transaction, i.e., t₂, in PWT is chosen to be modified. According the heuristic shown in Figure 3, the item '1' in t₂ are removed. Hence, ||SFI.t₂|| and ||SFI.t₃|| will be reduced by 1. SFI.t₃ is removed from SFI because the (||SFI.t₃|| / |D|) < min_support. The process is repeated until the SFI is empty. Finally, the FHSFI algorithm removes the item '1' in t₂, the item '6' or '8' in t₅ (select randomly), and the item '1' in t₁. Now all sensitive frequent itemsets in SFI have been hidden. ■

IV. PERFORMANCE EVALUATION

We have performed our experiments on a notebook with 1.5G MHz processor and 512 MB memory, under Windows XP operating system. The IBM data generator [11] is used to synthesize the databases for the experiments. Databases with sizes 5K, 10K, 15K, 20K, 25K, and 30K are generated for the series of experiments. The average length of transactions of each database is 10 and 50 items in the generated database. The minimum support threshold given is 30%. The experimental results are obtained by averaging from 5 independent trials with different SFIs.

The performance of the FHSFI algorithm has been measured according to three criteria: CPU time requirements, side effects produced, and the number of entries modified. Tables 8 and 9 present the experimental results for $|SFI|=5$ and $|SFI|=10$, respectively.

The CPU time requirements, side-effect evaluation, and the number of entries modified for varied $|D|$ and $|SFI|$ are shown in Figures 6, 7, and 8, respectively. Table 9. Experiment results for $|SFI|=10$

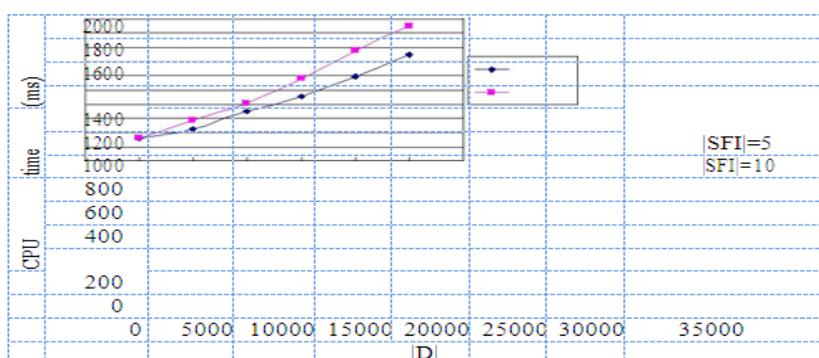


Figure 6. CPU time requirements

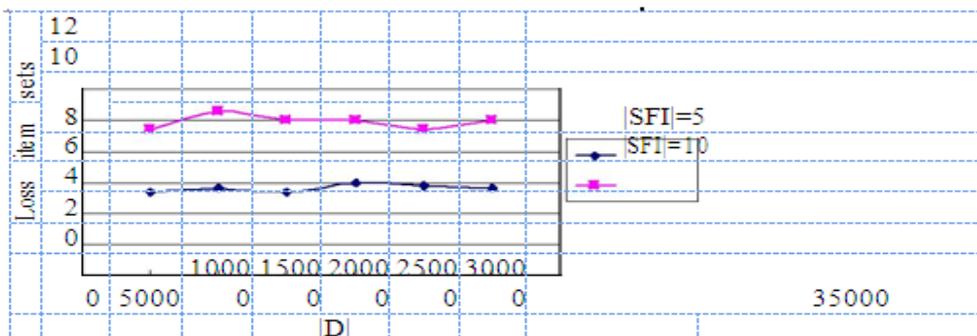
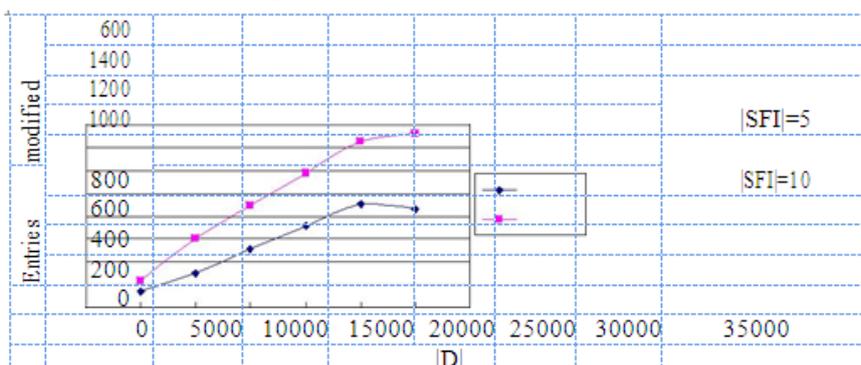


Figure 7. The side-effect evaluation



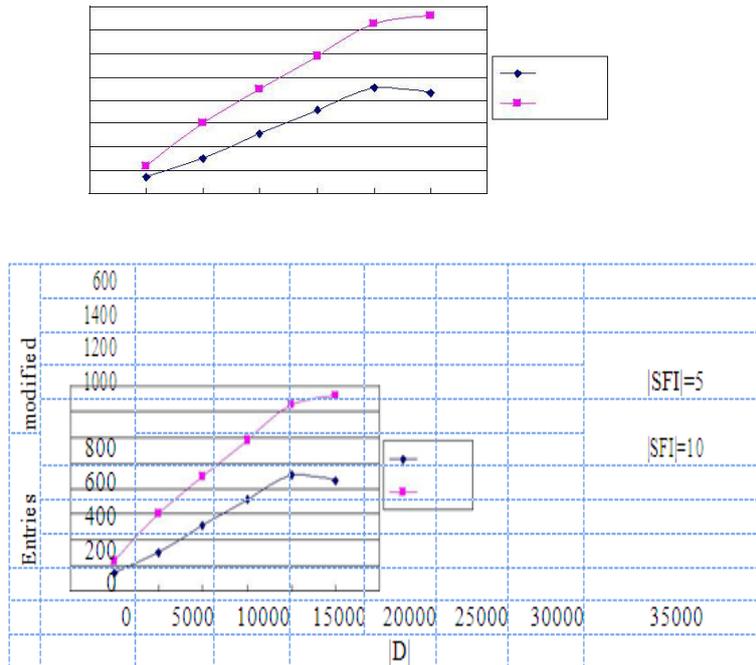


Figure 8. The number of entries modified
The experimental results for FHSFI can be summarized as follows:

- As shown in Figure 6, the CPU time is linear growth with the size of database and is scalable with the size of SFI.
- The number of loss itemsets is independent of the size of database, but linear-related with the size of SFI sets, which can be discovered in Figure 7.
- The number of the modified entries depends on the size of the database and the size of SFI. However, since the heuristic procedures are used to determine the order of modifications, we can observe in Figure 8 that only a small part of transactions in the database are modified. For $|D|=10000$, only 600 transactions are modified for completely hiding the 10 item sets in SFI.

V. CONCLUSIONS AND FURTHER WORK

In this paper, we have presented the FHSFI algorithm in order to fast hide sensitive frequent itemsets with limited side effects. The correlations between the sensitive itemsets and each transaction in the original database are analyzed. A heuristic function to obtain a prior weight for each transaction is given. The order of transactions to be modified can be efficiently decided by the weight for each transaction. This will reduce the time to deal with the transactions whose modification is not helpful for hiding the given sensitive frequent itemsets. In other words, the number of transactions in D that we have to deal with could also be reduced.

Our approach has achieved the following goals: 1) all SFI can be completely hidden while without generating all frequent itemsets; 2) limited side effects are generated; 3) any minimum support thresholds are allowed; and 4) only one database scan is required. In this research, one of our goals is hiding all SFI with limited side effects, but our algorithm still causes some loss rule sets. We are currently considering extensions on the algorithms to solve the problem. Another one is to apply the ideas introduced in this paper to fast hide sensitive association rules. These issues could be studied in the future.

REFERENCES

- [1] M. Atallah, E. Bertino, A. Elmagarmid, M. Ibrahim, V. Verykios, "Disclosure limitation of sensitive rules" *Knowledge and Data Engineering Exchange*, pp. 45-52, 1999.
- [2] Vassilios S. Verykios, A.K. Elmagarmid, E. Bertino, Y. Saygin, and E. Dasseni, "Association Rule Hiding," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 4, pp. 434-447, 2004.
- [3] Shyue-Liang Wang, "Hiding sensitive predictive association rules", *Systems, Man and Cybernetics, 2005 IEEE International Conference on Information Reuse and Integration*, vol. 1, pp. 164-169, 2005.
- [4] Ali Amiri, "Dare to share: Protecting sensitive knowledge with data sanitization", *Decision Support Systems archive* vol. 43, issue 1, pp. 181-191, 2007.
- [5] Yi-Hung Wu, Chia-Ming Chiang, and Arbee L.P. Chen, "Hiding Sensitive Association Rules with Limited Side Effects", *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, issue 1, pp. 29 - 42, 2007.

Authors Profile



Mr. Shaik.Mahammad Rafi –He was born in Rajampet, Kadapa, A.P, India in 1990.He is studying M.Tech in the department of Computer Science and Engineering at **Global College of Engineering and technology**, Kadapa. He has done Bachelor's of Technology from JNTUA University in the year 2011 in Information Technology.



Mr. M.SUMAN- He was born in Rajampet, Kadapa, A.P, India in 1985.He is Masters of Technology in Computer Science and Engineering from JNTUH University in the year 2013 at **Vathsalya Institute of Technology & Science**, Anantharam Bhongiri, Nalgonda . He has given guidance to many students in their thesis work of M.Tech. He has also contributed in the research work on Image Processing with his papers. He is presently working as Asst. Professor in **Mekapati Rajamohan Reddy Institute of Technology and Science** ,Udayagiri,SPSR Nellore. He has done Bachelor's of Technology from JNTUH University in the year 2006 in Computer Science & Engineering at **Annamacharya Institute of Technology & Science**, Rajampet, Kadapa.



Mr. MAJJARI SUDHAKAR- He was born in Raghunathapuram, Badvel, Kadapa, A.P, India in 1983.He is Masters of Technology in Information Technology from JNTUH University in the year 2010 at **Aurora's Scientific Technological & Research Academy** , Bandlaguda , Hyderabad.He has 3 years of Teaching Experience given guidance to many students in their thesis work of M.Tech. He has also contributed in the research work on Software Engineering with his papers. He is presently working as Asst. Professor in **Mekapati Rajamohan Reddy Institute of Technology and Science** ,Udayagiri,SPSR Nellore. He has done Bachelor's of Technology from JNTUH University in the year 2006 in Computer Science & Engineering at **Sri Venkateswara College of Engineering & Technology**,R.V.S Nagar Chittur.



Mr. P.RAMESH- He was born in Naidupet, S.P.S.R.Nellore, A.P, India in 1989.He is studying M.Tech in the department of Computer Science and Engineering at **SIR C.V.RAMAN INSTITUTE OF TECHNOLOGY & SCIENCE**, Tadipatri, Anantapur. He has done Bachelor's of Technology from JNTUA University in the year 2011 in Computer Science & Engineering



Mr. P.Venkata Ramanaiah –He was born in Khajipet,Kadapa,A.P,India. He is Master of Technology in Computer Science and Engineering at Madanapalle Institute of Technology and Science from JNTUA University in the year 2012. He has given guidance to many students in their thesis work of M.Tech. He has also contributed in the research work on Data Mining with his papers. He was presently working as Assistant. Professor in **Global College of Engineering and technology**,GCET, kadapa.

Effect of Controlling Parameters on Heat Transfer during Jet Array Impingement Cooling of a Hot Steel Plate

Purna C. Mishra¹, Santosh K. Nayak¹, Durga P. Ghosh¹, Manoj Ukamanal¹,
Swarup K. Nayak¹, Susant K. Sahu¹

¹ School of Mechanical Engineering, KIIT University, Bhubaneswar – 751024, Odisha, India

ABSTRACT:

This paper describes the experimental results on heat transfer characteristics of array of jet impingement cooling of a steel plate. The experiments were conducted on a stationary electrically heated steel plate. A commercially available shower was used to generate array of jets. The Time dependent temperature profiles were recorded by NI-cRIO DAS at the desired locations of the bottom surface of the plate embedded with K-type thermocouples. The controlling parameters considered in the experiments were water pressure, mass impingement density, mass flow rate, shower exit to surface distance respectively. Effects of these parameters on cooling rate were analysed through plots in the MS-EXCEL environments. The experimental results showed a dramatic improvement of heat transfer rate from the surface and the results established good optimal cooling strategies.

KEYWORDS: Array of Jets, Heat transfer coefficient, Heat transfer enhancement, Jet Impingement Cooling, Optimal cooling, Patternator, Statistics

Nomenclature:

ΔT	Measured temperature difference
ΔT^*	Non-dimensional temperature difference
T_c	Temperature of cooling water
T	Time
t^*	Non-dimensional time
l	Length of the steel plate
α	Thermal Diffusivity
CR	Cooling rate
CR*	Non-dimensional cooling rate
HTC	Heat Transfer Coefficient
D	Shower exit to surface distance
lt	Liter
min	Minute
m	Meter
s	Second
Pw	Water Pressure

I. INTRODUCTION:

Heat transfer enhancement is one the major parameters required to improve the performance of the thermal systems in many industries such as electronics, aerospace, automotive and steel manufacturing. Jet impingement cooling and spray impingement cooling are two of the most effective ways to improve the rate of heat transfer from the hot metal surface. Impingement cooling helps to achieve desired cooling rates from the surface by appropriate parametric control during the cooling process. Hence this process finds its use in many cooling application in particular to metal processing industry So far a great deal of experimental and computational work has been done to study these effects using single jet water impingement [1 – 4]. Lytle and Webb [5] have studied the effect of very low nozzle to plate spacing ($z/d < 1$) on local heat transfer distribution on a flat plate impinged by a circular jet issued by a long pipe nozzle which allows for fully developed flow at the nozzle exit. Gardon and Cobonpue [6] have reported the heat transfer distribution between circular jet and flat plate for the nozzle plate spacing greater than the two times the diameter of the jet, both for single jet and array of jets. Owen and Pulling [10] presented a model for the transient film boiling of water jets impinging on a hot metal surface of stainless steel and nimonic alloy. In their study, wetting of the surface was assumed to

occur when the temperature of the surface falls to the Leidenfrost temperature.. Kumagai et al. [11] conducted an experimental investigation of cooling a hot thick copper plate by impinging a plane water jet to clarify the transient behavior of boiling heat transfer performance along the surface and the temperature profile inside the body as well as on the surface. This experimental result indicated a time delay of approximately 100 s before the movement of the wetting front for a saturated jet with velocity 3.5 m/s and an initial solid temperature of 400 °C. One of the cooling means of achieving higher rates of heat transfer is using water jet impingement. It has been proven an effective means and widely applied in steel industries. Compared with the other cooling methods, it definitely is a rapid steel cooling method and reveals an outstanding advantage in improving heat transfer efficiency in cooling process of high temperature steel. According to Wolf, et al. [12], for single-phase convection, the heat transfer coefficient for water impingement cooling system exceeds 10 KW/m².

Now a days researchers are concentrating more on jet array impingement cooling rather than on single impinging jet, because array of jets maintains more consistent surface temperature and cools larger areas. Enhancement of heat transfer on an impinging surface by impinging an array of jets by minimizing a cross flow effect was studied by Nuntadusit et al. [13]. In his study conventional round orifices (of aspect ratio AR = 1) was substituted with elongated orifice (AR = 4 and 8). An experimental investigation of single phase and flow boiling heat transfer of submerged micro jet array was conducted with R134a by Eric A. Browne et al. [14]. Boiling experiments were performed with liquid subcooling of 10, 20 and 30°C for jet velocities of 4, 7 and 10 m/s. It was found that boiling enhances heat transfer with a maximum heat flux of 590 W/cm². Shyy Woei Chang et al. [15] presented heat transfer measurements by impingement of array of jets on surface roughened by concave dimples. They experimentally revealed the isolative and interactive influences of surface topology on local and spatially averaged heat transfers. S. Caliskan and S. Baskaya [16] presented heat transfer measurements over surface having V – shaped ribs (V - SR) and convergent – divergent ribs (CD - SR) by circular impinging jet array using thermal infrared camera. The effect of nozzle geometry in liquid jet impingement cooling was detailed by Brian P. Whelan and Antony J. Robinson [17]. In this paper effect of various controlling parameters on impingement cooling of a hot steel plate by an array of jets has been experimentally investigated. Commercially available water showers were used to generate array of jets and impinged on electrically heated square steel plate of side 120mm and thickness 4mm. The local temperatures are measured by using embedded K-type thermocouples. The temperature data were accessed by NI CRIO-DAS and used to investigate the heat transfer characteristics of jet cooling and characterize the showers to achieve the optimum cooling effect.

II. EXPERIMENTAL SET UP:

In the current paper the experimental set up is concentrated on rapid cooling of hot flat steel plate as shown in figure 1. The experimental set consists of heating arrangement, jet array impingement system and data acquisition system. Electrical coil heater of 2.5 KW, bounded over asbestos plate was used to heat up the selected test piece. To ensure uniform heating a proper gap was maintained between the lower surface of the test plate and the upper heater surface. For generating jet arrays, a commercially available shower having 80 holes on 50mm diameter periphery was used. The flow rate was regulated by changing the pressure at which the water was being delivered to the shower. The test piece was a square plate of side 120mm and 4mm thickness. It was embedded by four K – type thermocouples at suitable locations to record the temperature of the steel plate. The bed was designed in such a manner that the distance between the test piece and the shower can be varied by means of a hand wheel. A mechanical patternator was used to measure the impingement density. The patternator consists of an array of small collecting tubes arranged parallel to the main jet axis. NI CRIO – DAS aided with LABVIEW software was employed for acquiring the raw data, analyzing the data and presenting the results.

III. EXPERIMENTAL PROCEDURE:

The steel plate was heated to a temperature slightly higher than the test temperature to compensate for the heat loss during transfer of the steel plate from the heater to test bed (underneath the jet array system). When the temperature of the steel plate reached the required temperature the jet array was impinged on the hot steel surface and temperature - time data was generated with aid of NI - CRIO. The initial temperature of the steel plate was varied between 750°C to 950°C and the jet to surface distance was taken as 70mm, 135mm and 255mm. The mass flow rate was taken for different values between 0.5 to 9.56 lt./min. The mass flux density for different shower exit to surface gap was measured by means of a mechanical patternator. These were the same locations where the heat transfer experiments were subsequently conducted.

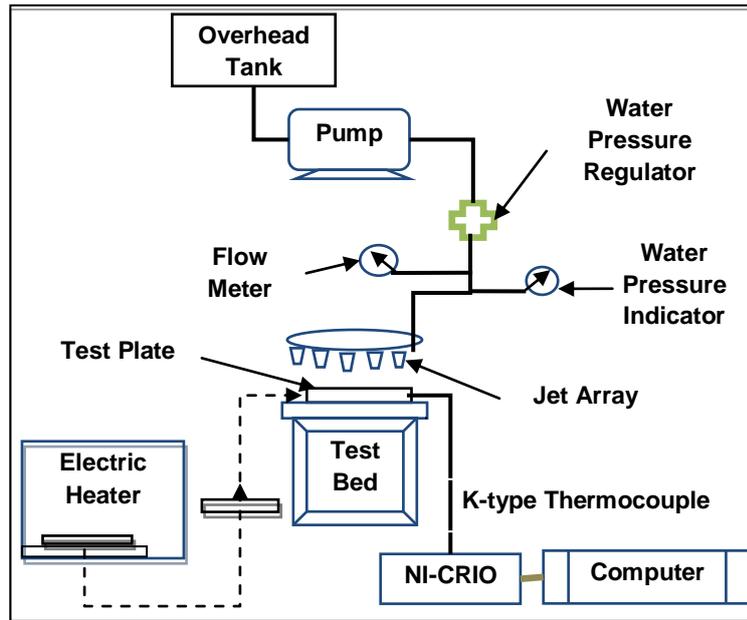


Figure 1. Schematic of experimental set up

IV. RESULTS AND DISCUSSION:

The variations of temperature with respect to time were measured at various key locations on the steel plate. From these data the rate of heat transfer from the plate were calculated. This paper mainly illustrates the relationships between the rate of heat transfer and the primary parameters: water pressure, mass flow rate, mass impingement density and shower exit to surface distance as well as cooling time.

4.1. Temperature – Time history:

Figure 2 shows the non-dimensional temperature difference against non-dimensional time for water pressure range of 2 - 4 bar. The non-dimensional temperature difference is taken as the ratio of the measured temperature difference to the cooling water temperature as shown in equation 1. The non-dimensional cooling time is calculated by using equation 2.

$$\Delta T^* = \Delta T / T_c \tag{1}$$

$$t^* = (l^2 / \alpha) * t \tag{2}$$

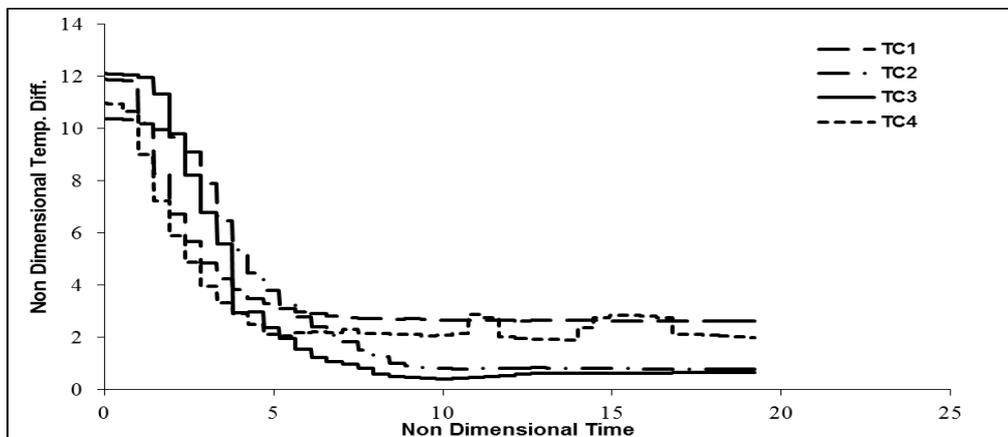


Figure 2. Non dimensional temperature difference vs non dimensional time for various thermocouple location

The temperature difference profile for different thermocouple locations in figure 2 indicates fairly uniform trend over the entire impinging surface. From the figure 3 shown below it is seen that as we increase the

pressure, the fall in non-dimensional temperature difference becomes steep. But when the pressure is increased from 3.5 bar to 4 bar, we observe the steepness reduces. This might be due to the fact that as we increase the pressure, the atomization of water particles increases and blow off occurs there by not contributing significantly to the cooling process.

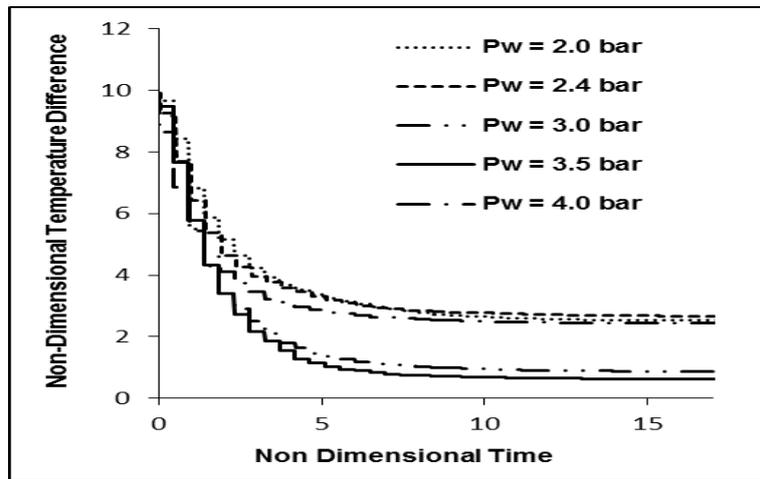


Figure 3. Non-dimensional temperature difference vs. non-dimensional time for different water pressure (pw)

4.2. Cooling rate:

The cooling rate is defined as the rate of change of temperature. The cooling rate is represented by equation 3.

$$CR = \Delta T / t \tag{3}$$

Figure 4 represents the variation of water flow rate with water pressure. Here we find that water flow rate varies directly with water pressure. Figure 5 plots the variation in non-dimensional cooling rate to that of water pressure for different jet exit to surface distance. The non-dimensional cooling rate is given by equation 4.

$$CR^* = \Delta T^* / t^* \tag{4}$$

Figure 5 clearly implies that as we increase the jet exit to surface distance the cooling rate reduces. This happens because when we increase the distance between the jet exit and the cooling surface, the cone angle increases and the amount of water impinging on the surface reduces.

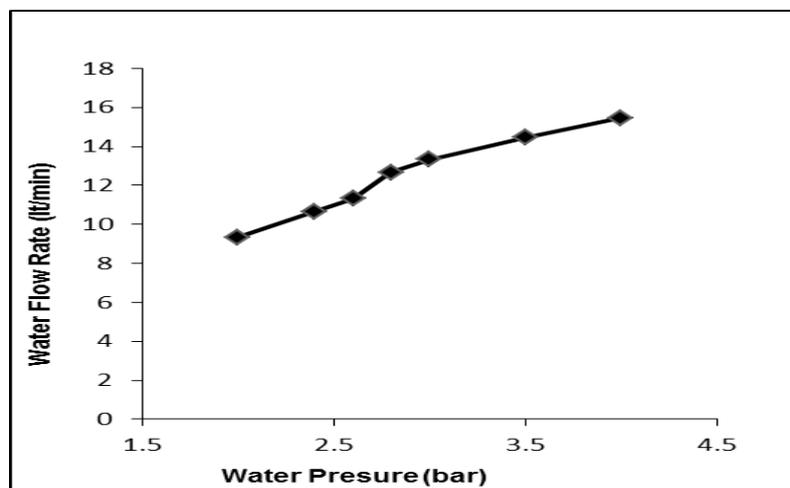


Figure 4. Water flow rate vs. water pressure

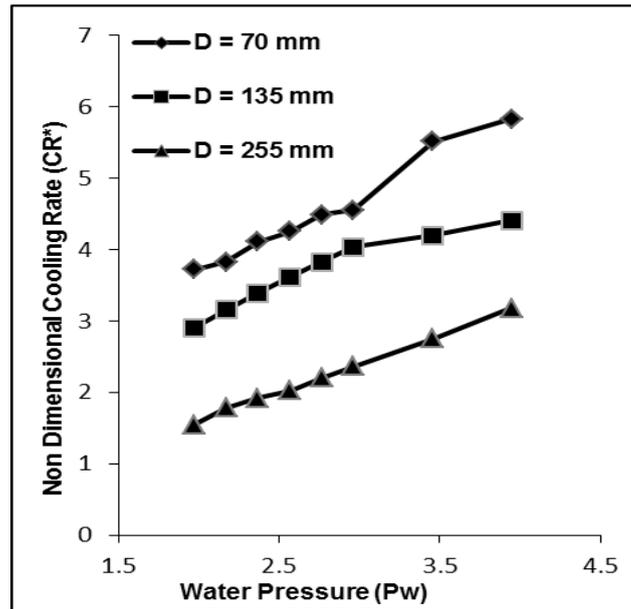


Figure 5. Non dimensional cooling rate vs. water pressure for d= 70mm, 135mm and 255mm

4.3. Mass Impingement Density:

A simple mechanical patternator was used to measure the mass impingement density. The behavior of cooling rate with respect to mass impingement is shown in figure 6.

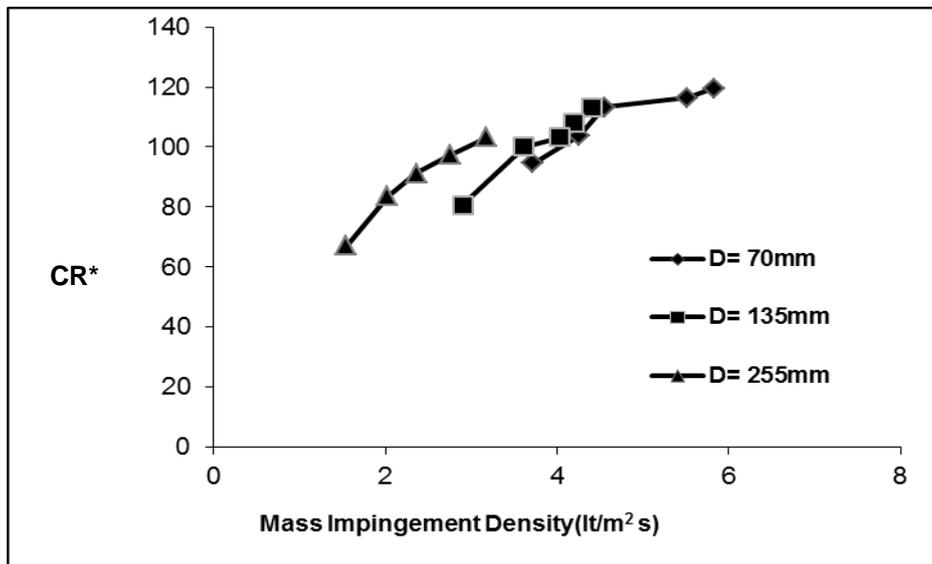


Figure 6. Non-dimensional cooling rate vs. mass impingement density for different shower exit to surface distance

From figure 6 it is pretty obvious that for D = 70mm the mass impingement density is more and hence the cooling rate increases.

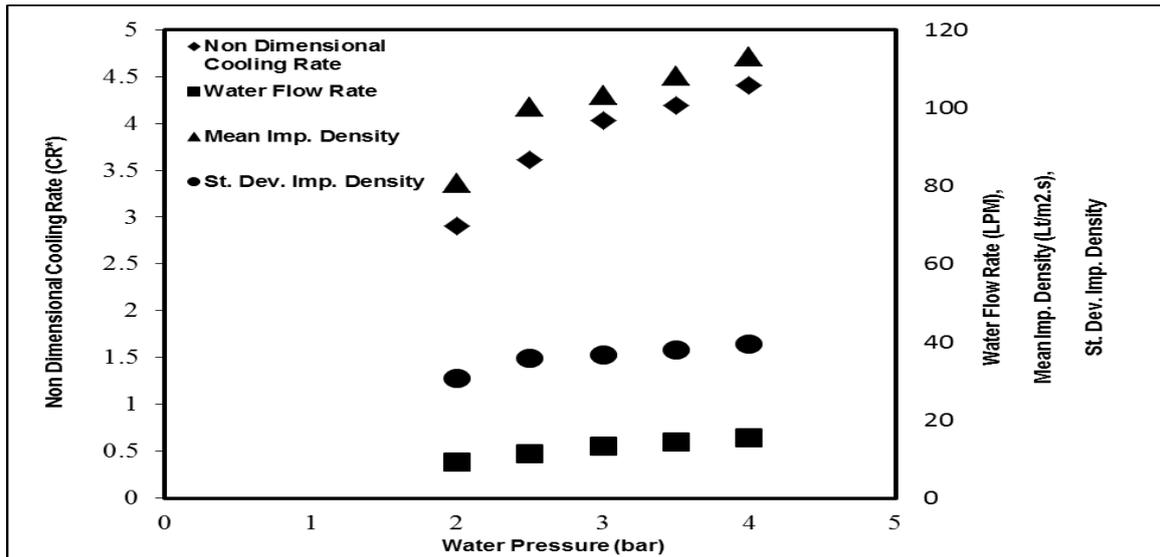


Figure 7. Variation of non-dimensional cooling rate with respect to water pressure, water flow rate, mean impingement density and standard deviation impingement density

It is observed from Figure 7 that with increase in water pressure, the water flow rate, impingement density and non-dimensional cooling rate increase monotonically.

4.4. Heat Transfer Coefficient:

The average heat transfer coefficient “h” was calculated using the cooling water temperature and the surface temperature of the hot steel plate. Figure 8 shows the variation of Heat transfer coefficient with non-dimensional temperature difference. The high value of Heat transfer coefficient indicates that the heat transfer taking place is predominantly convective heat transfer.

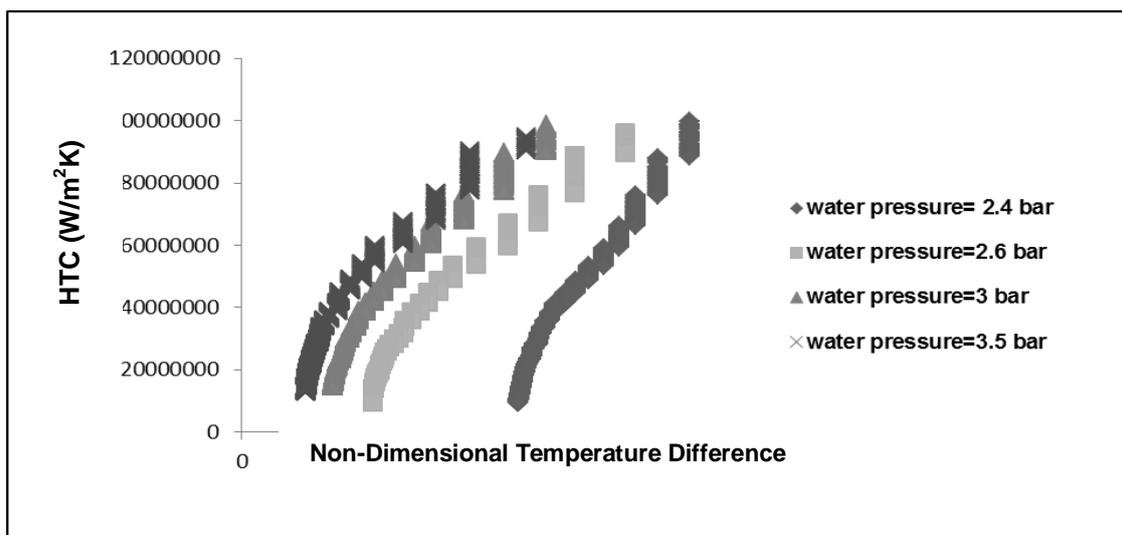


Figure 8. Heat transfer coefficient vs. non dimensional temperature difference for variations in water pressure.

V. CONCLUSION:

In the present study, the effects of various parameters such as water pressure, mass impingement density, mass flow rate and shower exit to surface distance on the heat transfer rate from a hot metal surface by impingement of an array of jets was experimentally studied. The major findings are as follows:-

- The cooling rate is uniform over the entire surface of the steel plate.

- The maximum drop in temperature difference with respect to time was recorded for a water pressure of 3.5 bar.
- At a shower exit to surface distance of 70mm the maximum cooling rate was obtained.
- With increase in mass impingement the cooling rate increased gradually.
- High value of heat transfer coefficient indicates active convective heat transfer.

From the above results it is can be said that to optimize the cooling rate the jet exit to surface distance and the water pressure are to be optimized.

REFERENCES

- [1] J.N.B Livingood, P. Hrycak. Impingement heat transfer from turbulent air jets to flat plates – a literature survey NASA Technical Memorandum (NASA TM X-2778) 1970.
- [2] H. Martin, Heat and mass transfer between impinging gas jets and solid surface. *Adv. Heat transfer* 13, 1 -60. 1977.
- [3] J.W. Baughn, S. Shimizu, Heat transfer from a surface from uniform heat flux and an impinging jet. *J. Heat Transfer* 111, 1096-1098, 1989.
- [4] R. Gardon, C. Akfirat, The role of turbulence in determining heat transfer characteristics of impinging jet. *Int. J. Heat Mass Transfer* 8, 1261-1272, 1965.
- [5] D. Lytle, B.W. Webb, Air jet impingement heat transfer at low nozzle plate spacings. *Int. J. Heat Mass Transfer* 37, 1687–1697, 1994.
- [6] R. Gardon, J. Cabonpue, Heat transfer between a flat plate and jets of air impinging on it. *Int. Develop. Heat Transfer, ASME*, 454-460, 1962.
- [7] R.G. Owen, D.J. Pulling, “Wetting delay: .lm boiling of water jets impinging hot .at metal surfaces”, in: T. Nejat Veziroglu (Ed.), *Multiphase Transport Fundamentals, Reactor Safety, Applications*, vol. 2, Clean Energy Research Institute, University of Miami, pp. 639–669, 1979.
- [8] S. Kumagai, S. Suzuki, Y. Sano, M. Kawazoe, “Transient cooling of a hot metal slab by an impingement jet with boiling heat transfer”, *ASME/JSME Thermal Eng. Conf.* 2, 1995.
- [9] D.H. Wolf, F.P. Incropera, R. Viskanta, *Jet Impingement Boiling*. *Advances in Heat Transfer*. Vol. 23, 1-131, 1993.
- [10] C. Nuntadusit, M. Wae-hayee, P. Tekasakul, S. Eiamsa-ard, “Local heat transfer characteristics of array impinging jets from elongated orifices”. *International Communications in Heat and Mass Transfer* 39, June 18, 1154–1164, 2012.
- [11] Eric A. Browne, Gregory J. Michna, Michael K. Jensen, Yoav. Peles., “Microjet array single-phase and flow boiling heat transfer with R134a”. *International Journal of Heat and Mass Transfer* 53, March 15, 5027–5034, 2010.
- [12] Shyy Woei. Chang, Shyr Fuu. Chiou, Shuen Fei. Chang, “Heat transfer of impinging jet array over concave-dimpled surface with applications to cooling of electronic chipsets”. *Experimental Thermal and Fluid Science* 31, June 17, 625–640, 2007.
- [13] S. Caliskan, S. Baskaya, “Experimental investigation of impinging jet array heat transfer from a surface with V-shaped and convergent-divergent ribs”. *International Journal of Thermal Sciences*, May 18, 59 234-246, 2012.
- [14] Brian P. Whelan, Anthony J. Robinson, “Nozzle geometry effects in liquid jet array impingement”. *Applied Thermal Engineering* 29 2211–2221 2009.

Design and Testing Of Prefix Adder for High Speed Application by Using Verilog HDL

¹, Rajitha Chandragiri, ², P. Venkata Lavanya

¹, Research Scholar, ECE Dept, TKR College of Engineering & Technology, Hyderabad, AP-INDIA

²Assoc Prof, ECE Dept, TKR College of Engineering & Technology, Hyderabad, AP-INDIA

ABSTRACT:

Parallel prefix adder is the most flexible and widely used for binary addition. Parallel Prefix adders are best suited for VLSI implementation. Numbers of parallel prefix adder structures have been proposed over the past years intended to optimize area, fan-out, and logic depth and inter connect count. This paper presents a new approach to redesign the basic operators used in parallel prefix architectures. The number of multiplexers contained in each Slice of an FPGA is considered here for the redesign of the basic operators used in parallel prefix tree. The experimental results indicate that the new approach of basic operators make some of the parallel prefix adder's architectures faster and area efficient.

Keywords: Ripple Carry Adder, Parallel Prefix, Kogge Stone, ASIC, Sparse Adders

I. INTRODUCTION

Addition is a timing critical operation in today's floating point units. In order to develop faster processing, an end-around carry (EAC) was proposed as a part of fused-multiply-add unit which performs multiplication followed by addition [5]. The proposed EAC adder was also investigated through other pre fix adders in FPGA technology as a complete adder [6]. In this thesis, we propose a 128-bit standalone adder with parallel prefix end around carry logic and conditional sum blocks to improve the critical path delay and provide flexibility to design with different adder architectures. In previous works, CLA logic was used for EAC logic. Using a modified structure of a parallel prefix $2n + 1$ adder provides flexibility to the design and decreases the length of the carry path. After the architecture is tested and verified, critical path is analyzed using FreePDK45nm library. Full custom design techniques are applied carefully during critical path optimization. Critical path analysis provides fast comparison of the total delay among different architectures without designing the whole circuit and a simpler approach to size the transistors for lowest delay possible. As a final step, datapath is designed as a recurring bitslice for fast layout entry. The results show that the proposed adder shows 142ps delay, 2.42mW average power dissipation, and 3,132 sq. micron area assuming there is not much routing area overhead in the estimated area.

Binary addition is the most fundamental and frequently used arithmetic operation. A lot of work on adder design has been done so far and much architecture have been proposed. When high operation speed is required, tree structures like parallel-prefix adders are used [1] - [10]. In [1], Sklansky proposed one of the earliest tree-prefix is used to compute intermediate signals. In the Brent-Kung approach [2], designed the computation graph for area-optimization. The KS architecture [3] is optimized for timing. The LF architecture [4], is proposed, where the fan-out of gates increased with the depth of the prefix computation tree. The HC adder architecture [5], is based on BK and KS is proposed. In [6], an algorithm for back-end design is proposed. The area minimization is done by using bitwise timing constraints [7]. In [8], which is targeted to minimize the total switching activities under bitwise timing constraints. The architecture [9], saves one logic level implementation and reduces the fan-out requirements of the design. A fast characterization process for Knowles adders is proposed using matrix representation [10].The Parallel Prefix addition is done in three steps.

which is shown in fig1. The fundamental generate and propagate signals are used to generate the carry input for each adder. Two different operators black and gray are used here. The aim of this paper is to propose a new approach for the basic operators and make use of these operators in various parallel prefix adders to evaluate their performance with newly redesigned operators The rest of the paper is organized as follows: In section II, some background information about Parallel Prefix architecture is given. New design approach of

basic operators is discussed in section III. Experimental results are presented in section IV. Conclusions are drawn in section V.

Several papers have attacked the problem of designing efficient diminished adders. The majority of them rely on the use of an inverted end around carry (IEAC) n -bit adder, which is an adder that accepts two n -bit operands and provides a sum increased by one compared to their integer sum if their integer addition does not result in a carry output. Although an IEAC adder can be implemented by using an integer adder in which its carry output is connected back to its carry input via an inverter, such a direct feedback is not a good solution. Since the carry output depends on the carry input, a direct connection between them forms a combinational loop that may lead to an unwanted race condition [21]. To this end, a number of custom solutions have been proposed for the design of efficient IEAC adders. Considering the diminished-1 representation for modulo $2n \mp 1$ addition, [4], [5] used an IEAC adder which is based on an integer adder along with an extra carry lookahead (CLA) unit. The CLA unit computes the carry output which is then inverted used as the carry input of the integer adder. Solutions that rely on a single carry computation unit have also been proposed. Zimmermann [22], [23] proposed IEAC adders that make use of a parallel-prefix carry computation unit along with an extra prefix level that handles the inverted end-around carry.

Although these architectures are faster than the carry lookahead ones proposed in [24], for sufficiently wide operands, they are slower than the corresponding parallel prefix integer adders because of the need for the extra prefix level. In [24], it has been shown that the recirculation of the inverted end around carry can be performed within the existing prefix levels, that is, in parallel with the carries' computation. In this way, the need of the extra prefix level is canceled and parallel-prefix IEAC adders are derived that can operate as fast as their integer counterparts, that is, they offer a logic depth of $\log_2 n$ prefix levels. Unfortunately, this level of performance requires significantly more area than the solutions of [22], [23], since a double parallel-prefix computation tree is required in several levels of the carry computation unit. For reducing the area complexity of the parallel-prefix solutions, select-prefix [25] and circular carry select [26] IEAC adders have been proposed. Unfortunately, both these proposals achieve a smaller operating speed than the parallel-prefix ones of [24]. Recently, very fast IEAC adders that use the Ling carry formulation of parallel-prefix addition [27] have appeared in [28], that also suffer from the requirement of a double parallel-prefix computation tree. Although a modulo $2n \mp 1$ adder that follows the $\delta n \mp 1$ -bit weighted representation can be designed following the principles of generic modulo adder design [29], specialized architectures for it have appeared in [30], [31]. However, it has been recently shown [32] that a weighted adder can be designed efficiently by using an IEAC one and a carry save adder (CSA) stage. As a result, improving the design for an IEAC adder would improve the weighted adder design as well.

III. PARALLEL-PREFIX ADDITION BASICS

The binary adder is the critical element in most digital circuit designs including digital signal processors (DSP) and microprocessor datapath units. As such, extensive research continues to be focused on improving the power delay performance of the adder. In VLSI implementations, parallel-prefix adders are known to have the best performance. Reconfigurable logic such as Field Programmable Gate Arrays (FPGAs) has been gaining in popularity in recent years because it offers improved performance in terms of speed and power over DSP-based and microprocessor-based solutions for many practical designs involving mobile DSP and telecommunications applications and a significant reduction in development time and cost over Application Specific Integrated Circuit (ASIC) designs. The power advantage is especially important with the growing popularity of mobile and portable electronics, which make extensive use of DSP functions. However, because of the structure of the configurable logic and routing resources in FPGAs, parallel-prefix adders will have a different performance than VLSI implementations.

In particular, most modern FPGAs employ a fast-carry chain which optimizes the carry path for the simple Ripple Carry Adder (RCA). In this paper, the practical issues involved in designing and implementing tree-based adders on FPGAs are. An efficient testing strategy for evaluating the performance of these adders is discussed. Several tree-based adder structures are implemented and characterized on a FPGA and compared with the Ripple Carry Adder (RCA) and the Carry Skip Adder (CSA). Finally, some conclusions and suggestions for improving FPGA designs to enable better tree-based adder performance are given.

IV. CARRY-TREE ADDER DESIGNS

Parallel-prefix adders, also known as carry-tree adders, pre-compute the propagate and generate signals [1]. These signals are variously combined using the fundamental carry operator (fco) [2].

$$(g_L, p_L) \circ (g_R, p_R) = (g_L + p_L \cdot g_R, p_L \cdot p_R) \quad (1)$$

Due to associative property of the fco, these operators can be combined in different ways to form various adder structures. For, example the four-bit carry-lookahead generator is given by:

$$c4 = (g4, p4) \circ [(g3, p3) \circ [(g2, p2) \circ (g1, p1)]] \quad (2)$$

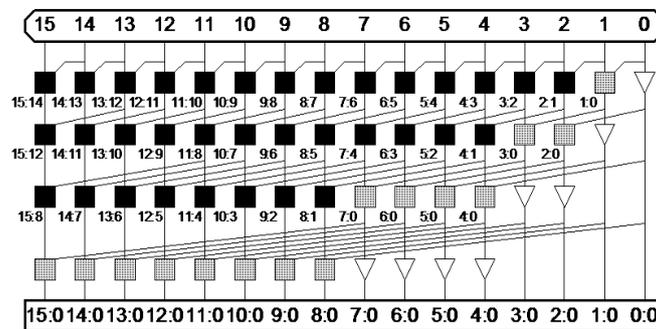
A simple rearrangement of the order of operations allows parallel operation, resulting in a more efficient tree structure for this four bit example:

$$c4 = [(g4, p4) \circ (g3, p3)] \circ [(g2, p2) \circ (g1, p1)] \quad (3)$$

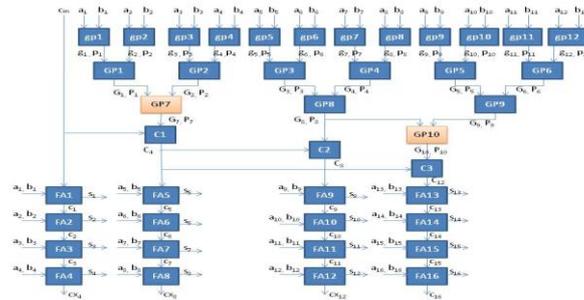
It is readily apparent that a key advantage of the tree structured adder is that the critical path due to the carry delay is on the order of $\log_2 N$ for an N-bit wide adder. The arrangement of the prefix network gives rise to various families of adders. For a discussion of the various carry-treestructures, see [1, 3]. For this study, the focus is on the Kogge-Stone adder [4], known for having minimal logic depth and fanout (see Fig 1(a)). Here we designate BC as the black cell which generates the ordered pair in equation (1); the gray cell (GC) generates the left signal only, following [1]. The interconnect area is known to be high, but for an FPGA with large routing overhead to begin with, this is not as important as in a VLSI implementation. The regularity of the Kogge-Stone prefix network has built in redundancy which has implications for fault-tolerant designs [5]. The sparse Kogge-Stone adder, shown in Fig 1(b), is also studied. This hybrid design completes the summation process with a 4 bit RCA allowing the carry prefix network to be simplified. Another carry-tree adder known as the spanning tree carry-lookahead (CLA) adder is also examined [6]. Like the sparse Kogge-Stone adder, this design terminates with a 4- bit RCA. As the FPGA uses a fast carry-chain for the RCA, it is interesting to compare the performance of this adder with the sparse Kogge-Stone and regular Kogge-Stone adders. Also of interest for the spanning-tree CLA is its testability features [7].

V. RELATED WORK

Xing and Yu noted that delay models and cost analysis for adder designs developed for VLSI technology do not map directly to FPGA designs [8]. They compared the design of the ripple carry adder with the carry-lookahead, carry-skip, and carry-select adders on the Xilinx 4000 series FPGAs. Only an optimized form of the carry-skip adder performed better than the ripple carry adder when the adder operands were above 56 bits. A study of adders implemented on the Xilinx Virtex II yielded similar results [9]. In [10], the authors considered several parallel prefix adders implemented on a Xilinx Virtex 5 FPGA. It is found that the simple RCA adder is superior to the parallel prefix designs because the RCA can take advantage of the fast carry chain on the FPGA. Kogge-Stone The Kogge-Stone tree [22] Figure 1.5 achieves both $\log_2 N$ stages and fan-out of 2 at each stage. This comes at the cost of long wires that must be routed between stages. The tree also contains more PG cells; while this may not impact the area if the adder layout is on a regular grid, it will increase power consumption. Despite these cost, Kogge-Stone adder is generally used for wide adders because it shows the lowest delay among other structures.



Another carry-tree adder known as the spanning tree carry-lookahead (CLA) adder is also examined [6]. Like the sparse Kogge-Stone adder, this design terminates with a 4-bit RCA. As the FPGA uses a fast carry-chain for the RCA, it is interesting to compare the performance of this adder with the sparse Kogge-Stone and regular Kogge-Stone adders. Also of interest for the spanning-tree CLA is its testability features [7].



This study focuses on carry-tree adders implemented on a Xilinx Spartan 3E FPGA. The distinctive contributions of this paper are two-fold. First, we consider tree-based adders and a hybrid form which combines a tree structure with a ripple-carry design. The Kogge-Stone adder is chosen as a representative of the former type and the sparse Kogge Stone and spanning tree adder are representative of the latter category. Second, this paper considers the practical issues involved in testing the adders and provides actual measurement data to compare with simulation results

VI. RESULTS

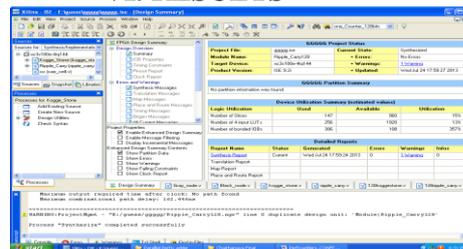


Fig.6.1 synthesis of 128 bit ripple carry adder

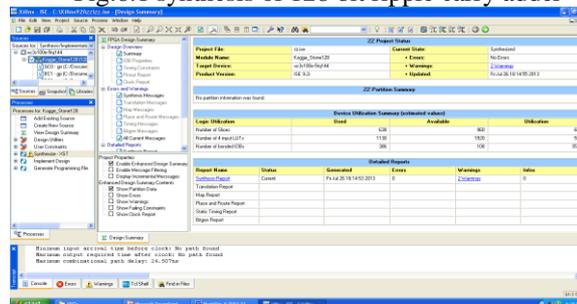


Fig.6.2 synthesis of 128 bit kogge stone adder

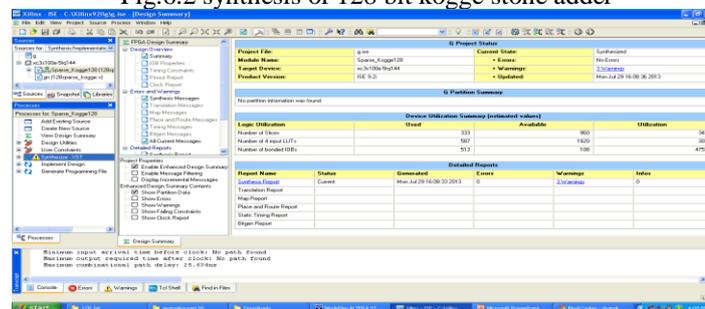


Fig.6.2 synthesis of 128 bit sparse-kogge adder

VII. CONCLUSIONS

This paper presents a new approach for the basic operators of parallel prefix tree adders. In KS, Sparse Kogge delay is reduced by this new approach, The same can be understood with reference to number of logic levels of implementation, as the logic levels are more delay increases. The area requirement can be considered from the utilization of LUTs, Slices and over all gate count. KS and Knowles adders occupies more area in new approach.

The KS adders are proven even faster than Sklansky adder but they occupy large area compared to Sklansky adder. Finally it can be concluded that the new approach for the redesign of basic operators will speed up the addition process of parallel prefix addition with some area overhead. The performance of these adders can be estimated for high bit-widths. This can be further used in Cryptographic applications, where the addition of more number of bits is necessary. The new approach for the parallel prefix adders can also be used to speed up the addition process in FIR filter and arithmetic operations like multipliers, etc.

REFERENCES

- [1] Kogge P, Stone H, "A parallel algorithm for the efficient solution of a general class of Recurrence relations", IEEE Trans. Computers, vol.C-22, No.8, pp 786-793, Aug.1973.
- [2] Brent R, Kung H, "A regular layout for parallel adders". IEEE Trans, computers, Vol.C-31, no.3, pp 260-264, March1982.
- [3] R. P. Brent and H. T. Kung, "A regular layout for parallel adders," IEEE Trans. Computers, vol. 31, no. 3, pp. 260–264, March 1982.
- [4] P. M. Kogge and H. S. Stone, "A parallel algorithm for the efficient solution of a general class of recurrence equations," IEEE Trans. Computers, vol. 22, no. 8, pp. 786–793, August 1973.
- [5] J. Sklansky, "Conditional sum addition logic," IRE Trans. Electron. Compute., vol. 9, no. 6, pp. 226–231, 1960.
- [6] J. Liu, S. Zhou, H. Zhu, and C.-K. Cheng, "An algorithmic approach for generic parallel adders," in ICCAD, November 2003, pp. 734–730.
- [7] R. Zimmermann, "Non-heuristic optimization and synthesis of parallel-prefix adders," in International Workshop on Logic and Architecture Synthesis, December 1996, pp. 123–132.
- [8] T. Matsunaga and Y. Matsunaga, "Timing-constrained area minimization algorithm for parallel prefix adders," IEICE Transactions on Fundamentals, vol. E90-A, no. 12, pp. 2770–2777, December 2007.
- [9] T. Matsunaga, S. Kimura, and Y. Matsunaga, "Power-conscious syntheses of parallel prefix adders under bitwise timing constraints," in Proc. the Workshop on Synthesis And System Integration of Mixed Information technologies(SASIMI), Sapporo, Japan, October 2007, pp. 7–14.

Risk Assessment of Bot Projects

¹, Sharmila Mane, ², Dr. S.S. Pimplikar

¹ Research Postgraduate ME (Construction & Management) MIT – Pune

² HOD – Civil Engineering Department MIT - Pune

ABSTRACT

The growth of the infrastructure sector in India has been relatively slow compared with the industrial and manufacturing sectors. The energy shortage, an inadequate transportation network, and an insufficient water supply system have caused a bottleneck in the country's economic growth. The Build-Operate-Transfer (BOT) scheme is now becoming one of the prevailing ways for infrastructure development in India to meet the needs of India's future economic growth and development. There are tremendous opportunities for foreign investors. However, undertaking infrastructure business in India involves many risks and problems that are due mainly to differences in legal systems, market conditions and culture. It is crucial for foreign investors to identify and manage the critical risks associated with investments in India's BOT infrastructure projects. Based on the survey, the following critical risks are identified: delay in approval, change in law, cost overrun, dispatch constraint, land acquisition and compensation, enforceability of contracts, construction schedule, financial closing, tariff adjustment, and environmental risk. The measures for mitigating each of these risks are also discussed. Finally a risk management framework for India's BOT infrastructure projects is developed. Main purpose of this paper is to investigate critical risks associated with Build Operate Transfer projects in India.

KEYWORDS: Risk management, BOT, Infrastructure projects, Mitigating measures, equity risks

I. SCOPE OF RISK ANALYSIS ON BOT PROJECTS

The long lasting implementation of project risk management in India can best be evidenced by the construction project procedure that has been in use for over 4 decades in India. The feasibility study was formally introduced into the procedure in 1992. A capital project (including infrastructure projects) must follow the procedure. When an organization has identified its need for a new facility, it must submit a project proposal defining the purpose, requirements and general aspects of the project, such as location, performance criteria, scope, layout, equipment, services and other requirements. The definition and planning of the project shall be carried out in coordination with agencies in charge such as provincial, municipal, autonomous region governments, central ministries or commissions. The project proposals of a medium or large sized project must be submitted to the agencies in charge for review and comments. The priority projects are subject to review and approval by the State Council.

The review and approval procedure makes sure that the project complies with the national economic and social development programs and there are sufficient resources available to the project. Once the proposal is approved site selection and feasibility study shall follow. The feasibility study involves the process of risk identification and analysis. Various matters should be considered when selecting the site for the proposed project and feasibility study is made, such as climate, topographical and geological conditions, resources, transportation, potential natural calamities, environment conservation, available services, and utilities and so on. Usually, several alternative sites and proposals should be considered and compared with each other in terms of the various factors influencing the project.

II. OBJECTIVES OF RISK ASSESMENT

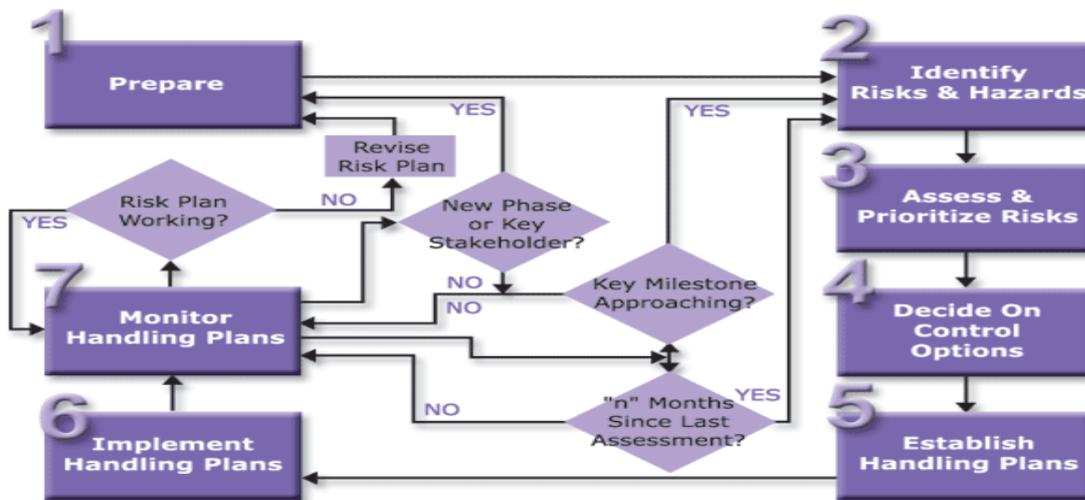
The experience of private investment in infrastructure in India over past years indicates that risks and pitfalls go together with opportunities. Proper identification, therefore, of the risks associated with investment in infrastructure in India and planning for effective responses thereto are essential for the private investors to be successful. In general, in order to be successful all capital projects shall meet the criteria and have the characteristics as listed below.

1. A credit risk rather than an equity risk is involved.
2. A satisfactory feasibility study and financial plan have been prepared.
3. The cost of product or raw material to be used by the project is assured.
4. A supply of energy at reasonable cost has been assured.
5. A market exists for the product, commodity or service to be produced.
6. The best way to appreciate the concerns of investors in infrastructure in India is to review and consider some of the common causes for their failures as shown below :

- Delay in completion, with consequential increase in the interest expense on construction financing and delay in the contemplated revenue flow.
- Capital cost overrun.
- Technical failure.
- Financial failure of the contractor.
- Government interference, inactions.
- Uninsured casualty losses.
- Increased price or shortages of raw materials.
- Technical obsolescence of the plant.
- Loss of competitive position in the market place.
- Expropriation.
- Poor management
- Financial insolvency of the host government.

In particular, for private investors to be successful in their infrastructure projects, these risks must be properly considered, monitored and avoided throughout the life of the projects

III. PICTORIAL REPRESENTATION OF RISK MANAGEMENT



IV. TYPES OF RISKS IN BOT PROJECT

The followings are the main findings through the literature review:

- a) The BOT scheme to financing infrastructure projects has many potential advantages and is a viable alternative to the traditional approach using sovereign borrowings or budgetary resources.
- b) BOT projects involve a number of elements, such as host government, the Project Company, lenders, contractors, suppliers, purchasers and so on. All of which must come together for a successful project.
- c) The application of the BOT scheme in Indian infrastructure development is being carried out stage by stage.
- d) There are two broad categories of risk for BOT projects: country risks and specific project risks. The former associated with the political, economic and legal environment and over which the project sponsors have little or no control. The later to some extent could be controllable by the project sponsors.
- e) A few researches of risk management associated with India's BOT projects focused on a particular sector. Different researchers appear to have different points of view on risk identification because they have approached the topic from different angles.

- f) A particular risk should be borne by the party most suited to deal with it, in terms of control or influence and costs, but it has never been easy to obtain an optimal allocation of risks. Risk management is a critical success factor of BOT projects.

Above points are as per KPMG report on India Infrastructure sector dated 22nd July 2010.

V. TYPES OF RISKS IN BOT PROJECT

Progress of a project is corresponding with the occurrence of risks. Risks have been categorized into three major captions; financing, political and technical risks. The successes of a project are measured by the overall project cost, duration and quality of the final product or services delivered. Usually the risks are corresponding with these three parameters

A. Financial Risk

1) Currency Risk

The investor or lenders is aware the existence of currency risk in any BOT projects and it does occur due to funding from international banks or foreign companies; creates volatility of the exchange rates. Bing et al. have stated that fluctuations in currency considered as an austere problem in international transactions.

2) Interest Rate Risk

In contrast, interest rate will affect the project in terms of borrowing and debt payments. Any fluctuation in the interest rate will definitely affect the lenders. An appropriate interest rate should be agreed upon the project. The lenders have to pay extra cost if the interest rate is far high or benefit them if the interest rate is low. More foreign investors or private sector could be attracted by providing interest rate guarantee by the host government in a BOT project. This approach has been adopted in Indonesian BOT toll road whereby the government has guaranteed on maximum interest rate, minimum revenue guarantee, debt guarantee, tariff guarantee and minimum tariff guarantee.

3) Equity Risk

Performance of the concessionaire is crucial in seeking for fund to implement a BOT project. Usually, equity risk is related with the performance of the company which is measured by the share price of the company. The higher the share price goes, in definite, benefit the shareholder but the lesser it goes will affect the prestige of the concessionaire. Capability of the company in raising capital for the BOT project is reflected on the share price. It has been believed that, the equity investors and other long term investors will only agree to provide the amount of funding for BOT project upon the promoter has proven their financial capability of the project over its entire lifespan. It is very difficult to attract domestic capital of debt and equity especially in East Asia when it is involving huge amount of investment in an infrastructure project. Nevertheless, the competence in carrying out detailed and comprehensive feasibility study, economic and risk assessment study, ensure the promoter to be in better position in obtaining domestic equity finance for funding the BOT project.

4) Foreign Exchange Risk

Fluctuations in foreign exchange are considered another major risk which might affect the BOT project during the construction and operation. Foreign companies who are interested to invest in another country should aware of the opportunities and threats associated with international currency transactions before they proceed. The Malaysian government has managed to reduce the foreign exchange risk by providing guarantees. The said guarantees was to absorb the shortfall when (1) the adverse exchange rate movements exceeds 15 percent on its offshore debt (2) adverse interest rate movements exceeds 20 percent on its floating rate offshore debt [2,7]. This approach was adopted in North-South Highway project and it benefited the promoter.

5) Commercial Risk

Commercial risk is described as a risk that can jeopardize the financial performance to the project. In spite of that, commercial risk in BOT project is characterized differently; Merna and Njiru have classified into three categories, risks related to the completion, during operation and risks related to input or output of the project [2]. Supply and off-take agreement between the supplier and the government is very crucial in mitigating the risk. In the agreement, the related parties will agree upon the required amount of input for instance coal in power plant project and the output generation is the electricity. This will allow the supplier to stock the materials upfront at lower price and the promoter able to generate the required output within the stipulated cost, not burdening the public by increasing the tariff. Apart from that, it will secure long term revenue for the promoter by selling the output to the client.

6) Liquidity Risk

Most of the BOT project, the revenues are generated from the operation. To ensure the success of the BOT project, it should be able to generate sufficient amount of revenue to settle the debt within the stipulated time frame. An amount of profit that can be generated from the operating facility is determined by conducting analysis on the projected revenue during operational phase. The failure to generate the required revenue will cause to liquidity risk.

7) Counterparty Risk

Inadequate support in terms of financial from the lender at specific time is defined as counterparty risk. It also can be interpreted as a credit risk. According to Lam and Chow, credit risk as the risk that the counterparty (partner of the joint venture) to any financial transaction is not being able to fulfill its commitment on the due date [16]. The debt capacity of the promoter will reduce when the credit risk arises. In a concession contract, transactions between two or more parties contain a risk that one party will default on an obligation of the commitment. Failure in financing the required cash flow for the BOT project is the most common issue that arises.

8) Economic Risk

This risk mostly related to the facility's operation which consist of materials supply, labour supply, equipment availability, inflations, tariffs, fiscal policies and exchange rates [17]. Project cash flow is affected by any financial aspects that relate to the economic parameters. Increment in supplier to stock the materials upfront at lower price and the promoter able to generate the required output within the stipulated cost, not burdening the public by increasing the tariff. Apart from that, it will secure long term revenue for the promoter by selling the output to the client

B. Political Risk

1) Sovereign Risk

Sovereign risk is a risks related to the provision of loans to foreign government and commonly used in banking world [3]. The risk is governed by the political environment of the country where the investment will take place, specifically, the location of BOT project commence. Sovereign risk occurs when the political environment is unstable and will affect the investor or promoter of the project. For instance in East Asia, some of the BOT project faced difficulties due to political instability such as in Thailand due to frequent change in political leaders. Apart from that, countries which are governed by different ideologies such as Libya and Saudi Arabia are also facing sovereign risk. A BOT project might face serious risk when there are changes in government's policy and regulations due to changes in ruling government as can be learned from past experience. The changes in bureaucracy level due to reshuffling might posses impact on the decision making process in a concession contract. By providing some kind of guarantee by the host government, the risk shall be prevented. In addition to that, the concession contract agreement should be based on the international order systems to safeguard the promoter.

2) Country Risk

Country risk is totally different from the sovereign risk. It related to overall investment climate in a specific country. The aspects that can contribute to country risk are socio-economic condition, internal or external conflicts that inflicting the country corruptions, ethnic tensions, policy and legal aspects. Before any BOT project implementation, the promoter should necessary conduct a thorough country risk profile and budgetary practices by a reliable third party (reputable management consultant or a good political analyst or both) to minimize the risk. Decision could be made based on the study and they should also seek assistance by referring to World Bank or Asian Development Bank. Nowadays, in East Asia, consortiums are established to undertake most of the BOT project. These consortiums consist of Engineering Procurement and Construction, (EPC) Contractors, Operation and Maintenance, (O&M) contractors and will be responsible for the feasibility study till implementation of the whole project and operation for a stipulated period of time. Every foreign investment is subjected to country risk due to unstable government and its component, and inadequate foreign reserves.

C. Technical Risk

Technical risk could be classified into construction risk and O&M risk. Essentially, technical risk is the most common and well understood form of risk. Technical risk is the subject of close surveillance. To minimize the technical risk, the concessionaire is responsible to evaluate the risk in detail to ensure the project will be constructed accordance to the design specification and host government's requirements and functioning well. Thus, well reputed and established consultant together with an experience contractor should be hired to implement the BOT project without any tolerance to the standard codes and practice.

1) Construction Risk

Unknown ground conditions, delay in procuring of construction materials, and price escalation of raw materials for construction such as an increase in the price of steel, copper or aluminum are the problems related to construction risk which occur during construction phase. In addition to that, poor design report, prolong construction schedule and changes in factor of production also contribute to construction risk. In the North-South Highway project in Malaysia has caused the promoter to bear the increment in the project cost. The initial estimated cost was US\$1.2 billion but due to hassles encountered during construction phase such as land acquisition problem and poor road design, the cost escalated to US\$1.8 billion [2,7,18]. The increment was almost fifty percent higher than estimated cost, this scenario was not unusual in East Asia since the local contractors are still learning and relying on the expertise of foreign based contractors which can be costly in terms of consulting fees. It is essential to make available the design report and to be vetted by the owner and consultant before any BOT project commences. Preferably, to have a third independent party to audit and comment on the design and construction methodology which would help to minimize the construction risk.

2) Operational and Maintenance Risk

During this phase there are several associated risks. One of them is when the performance of the facility is not to the required level due to technical problems. Selection of inefficient machineries and equipments during the implementation phase and poor workmanship during installation phase could cause the poor performance. The spare parts for the selected machinery and equipments for the facility are to be ensured that they are easily available at affordable cost. Throughout the concession period, the machinery and equipments will undergo some routine service due to wear and tear process, to optimize the performance. New available technology should be incorporated to ease the operation phase. Sometimes, initial cost is very high but in the long run it will benefit the consortium. The operation and maintenance team requires specialized technical skills and abilities in operating the facility. Inefficient team would lead to unnecessarily high cost of operating and resulting lesser revenue to the consortium. It is very important that a proper agreement should be established to ensure the interest of the operator is secured. The efficiency of the facility's operation could be increased by providing a maintenance manual and update it on a regular basis together with standard operating procedure.

1. RISK RESPONSE STRATEGIES

India's government officials and project managers use the risk response strategies that are available to them.

➤ Avoidance :

Risk avoidance is changing the project plan to eliminate the risk or condition or to protect the project objectives from its impact. Some risk events that arise early in the project can be dealt with by clarifying requirements, obtaining information, improving communication, or acquiring expertise. Reducing scope to avoid high-risk activities, adding resources or time, adopting a familiar approach instead of an innovative one, or avoiding an unfamiliar subcontractor may be examples of avoidance.

➤ Transference :

Risk transfer is seeking to shift the consequence of a risk to a third party together with ownership of the response. Transferring the risk simply gives another party responsibility for its management; it does not eliminate it. Transferring liability for risk is most effective in dealing with financial risk exposure. Risk transfer nearly always involves payment of a risk premium to the party taking on the risk. It includes the use of insurance, performance bonds, warranties, and guarantees. Contracts may be used to transfer liability for specified risks to another party. Use of a fixed-price contract may transfer risk to the seller if the project's design is stable. Although a cost-reimbursable contract leaves more of the risk with the customer or sponsor, it may help reduce cost if there are mid-project changes.

➤ Mitigation :

Mitigation seeks to reduce the probability and/or consequences of an adverse risk event to an acceptable threshold. Taking early action to reduce the probability of a risk's occurring or its impact on the project is more effective than trying to repair the consequences after it has occurred. Mitigation costs should be appropriate, given the likely probability of the risk and its consequences. Risk mitigation may take the form of implementing a new course of action that will reduce the problem—e.g., adopting less complex processes, conducting more seismic or engineering tests, or choosing a more stable seller. It may involve changing conditions so that the probability of the risk occurring is reduced—e.g., adding resources or time to the schedule. It may require prototype development to reduce the risk of scaling up from a bench-scale model. Where it is not possible to reduce probability, a mitigation response might address the risk impact by targeting linkages that determine the severity. For example, designing redundancy into a subsystem may reduce the impact that results from a failure of the original component.

➤ **Acceptance :**

This technique indicates that the project team has decided not to change the project plan to deal with a risk or is unable to identify any other suitable response strategy. Active acceptance may include developing a contingency plan to execute, should a risk occur. Passive acceptance requires no action, leaving the project team to deal with the risks as they occur. A *contingency plan* is applied to identified risks that arise during the project. Developing a contingency plan in advance can greatly reduce the cost of an action should the risk occur. Risk triggers, such as missing intermediate milestones, should be defined and tracked. A *fallback plan* is developed if the risk has a high impact, or if the selected strategy may not be fully effective. This might include allocation of a contingency amount, development of alternative options, or changing project scope. The most usual risk acceptance response is to establish a *contingency allowance*, or reserve, including amounts of time, money, or resources to account for known risks. The allowance should be determined by the impacts, computed at an acceptable level of risk exposure, for the risks that have been accepted.

VI. METHODOLOGY OF STUDY

Procedure :

This research study employed a combination of methods for an integrated qualitative and quantitative research methodology. The first stage was a comprehensive literature review together with lessons learned from the practice of BOT projects in developing countries, especially in India, to develop an initial list of risks associated with India's BOT infrastructure projects. In the second stage of instrument development, only the critical risks associated with India's BOT Infrastructure projects were chosen for study.

RISK MANAGEMENT FRAMEWORK FOR BOT INFRASTRUCTURE PROJECT

Based on the survey results and analysis as well as case studies, a risk management framework for investing in India's future BOT infrastructure projects can be proposed as follows.

Step 1: List all risks associated with the proposed BOT infrastructure project and then analyze these risks in order of importance. The more critical the risk, the more attention should be paid to it.

Step 2: For each risk, list corresponding mitigation measures as more as possible, and then examine the availability of mitigating measures in sequence based on their effectiveness. The more effective the measure, the higher the priority for adoption. Sometimes, a combination of several mitigating measures is needed to be adopted.

Step 3: For each risk and its mitigating measures, negotiate with Indian government and related entities to incorporate the risk mitigation measures, and fine tune the concession agreement and other agreements as much as possible to ensure that all of these risks are adequately covered.

Step 4: Allocate risks to related parties according to the principle that risk should be borne by the party most capable of controlling it. An optimal allocation of risks depends on the relative bargaining power of the parties and the potentiality of reward for taking the risks

Step 5: Adopt the risk allocation and security structure and enter into financing process for the project.

VII. CONCLUSION

In this research, the critical risks associated with India's BOT projects were investigated. The main conclusions are as follows:

- The identified critical risks in order of importance are: delay in approval, change in law, cost overrun, dispatch constraint, land acquisition and compensation, enforceability of contracts, construction schedule, financial closing, tariff adjustment, and environmental risk.
- The measures for mitigating each of these risks have been evaluated by respondents. Most of the measures were regarded as effective to some degree, however the most effective measures to mitigate each risk are:

- 1) For delay in approval, maintaining a good relationship with government authorities, especially officers at the state or provincial level;
- 2) For change in law, obtaining government's guarantees via adjusting either the tariff or extending concession period;
- 3) For cost overrun, entering into contracts with the project participants so that all share the responsibility and the incentive;
- 4) For dispatch constraint, entering into take-or-pay contracts with other parties;
- 5) For land acquisition and compensation, obtain government's guarantees to achieve timely acquisition of land;
- 6) For enforceability of contracts, making a credit judgment on the financial ability and Integrity of the contracting party to live up to its contractual obligation;
- 7) For construction schedule, choosing quality, trust-worthy Indian partners with knowledge of how to handle everyday construction issues;
- 8) For financial closing, equity financing and cooperation with government partners;
- 9) For tariff adjustment, negotiating to separate and redefine the tariff burden so that while some portions of the total tariff burden remained fixed other portions were either adjusted, re-scheduled or paid in foreign currency; and
- 10) For environmental risk, creating appropriate lines of communication and contacts with government authorities and agencies.

The risk management framework proposed by this research project will be easier to apply than others. It incorporates the findings from this research and provides step-by-step guidelines for foreign companies who intend to invest in India's infrastructure projects in the future. It also has the potential to help national, provincial, and city government to examine their approach to and services in support of BOT infrastructure projects. It suggests that mechanisms be reviewed to improve the communication and coordination links between different levels of government, that thought be given to developing mechanisms to coordinate actions by different government agencies and that the lessons learned from individual BOT projects be shared among government servants so that unintended barriers to BOT are dismantled.

REFERENCES

- [1]. Bond, Gary, and Laurence Carter, "Financing Private Infrastructure Projects: Emerging Trends from IFC's Experience," IFC Discussion Paper No. 23 (Washington: The World Bank, 1994).
- [2]. Grey, S. (1995), *Practical Risk Assessment for Project Management*, John Wiley & Sons Ltd, Chichester.
- [3]. International Finance Corporation, *Financing Private Infrastructure, Lessons of Experience No. 4* (Washington: The World Bank, 1996).
- [4]. Kleimeier, S., Megginson, W. L., 1998. A comparison of project finance in Asia and the West. In: Lang, L. H. P. (Ed.), *Project Finance in Asia. Advances in Finance, Investment and Banking*. Vol. 6. Elsevier - North Holland, Amsterdam, pp. 57-90.
- [5]. Risk Management in PPP Projects – IL&FS Report (*Construction Risk Management Conference India, August 2010*)
- [6]. Qiao, L., Wang, S. Q., Tiong, L. K. R. and Chan, T. S. (2001). Framework for Critical Success Factors of BOT Projects in China, *The Journal of Structured and Project Finance*, 7(1): 53-61.
- [7]. Wang, S. Q., Dulaimi, M. F. and Aguria, M. Y. (2002). *Building the External Wing of Construction: Managing Risk in International Construction Project*, Research Report, National University of Singapore.
- [8]. Wang, G. Q., Jia, X. L. (2005). Risk Management on the BOT Investment and Financing Mode, *India Water & Wastewater*, 21(9): 85-88
- [9]. World Bank. *World Development Report 1994-Infrastructure for Development*. Oxford University Press, Inc. New York, 1994

Image Fusion For Medical Image Retrieval

Deepali Sale¹, Dr. Madhuri Joshi², Varsha Patil³

Pallavi Sonare⁴, Chaya Jadhav⁵

^{1,3,4,5}Pad. Dr. D.Y. Patil Institute of Engineering and Technology, Pimpri, Pune, India

²College of Engineering, Pune-411005, India

ABSTRACT:

In medical imaging, various modalities provide different features of the human body because they use different physical principles of imaging. CT and MRI images with high spatial resolution provide the anatomical details, while PET and SPECT show the biochemical and physiological information but their spatial resolutions are not good enough. So it is very useful and important to combine images from multi-modality scanning such that the resulting image can provide both functional and anatomical information with high spatial resolution. In this paper we present a wavelet-based image fusion algorithm. The images to be fused are firstly decomposed into high frequency and low frequency bands. We select four groups of images to simulate, and compare our simulation results with the pixel addition, weighted averaging method and wavelet method based on min-max and subtraction based fusion rule. Then, the low and high frequency components are combined by using different fusion rules. Finally, the fused image is constructed by inverse wavelet transform. The various objective and subjective evaluation metrics and Quality are calculated to compare the results. The wavelet based fusion methods using different fusion rules is compared both subjectively as well as objectively. The experimental results show that the pixel minimum method is giving the better results in respect of MSE, SNR and using edge based quality metrics addition method observed to be better in preserving the edge information. One Image fusion method can be perfect for one particular application but may not for another application. So it depends on which information to extract, enhance, and reconstruct or retrieve to use the particular fusion method.

KEYWORDS: *Computed Tomography (CT), DWT (Discrete Wavelet Transform), Image fusion, MRI (Magnetic Resonance imaging), Quality.*

I. INTRODUCTION

In health care domain there are many research projects. The medical system is under many pressuring factors to offer the highest services, with efficiency, in the conditions of a growing population. To offer a real support for diagnosis, images have to be processed with different algorithms, for a better accuracy. A fusion based on transforms has some advantages over other simple methods, like: energy compaction, larger SNR, reduced features, etc. The transform coefficients are representative for image pixels. Clinical investigations are based more and more on medical imaging and physicians are faced very often with the difficulty of integrating the great amount of data. The specialists have to visualize and compare images from different medical modalities and to correlate the observed information with the clinical and auxiliary data. A fusion of multimodal images can be very useful for clinical applications such as diagnosis, modeling of the human body or treatment planning. Fusion of images taken at different resolutions, intensity and by different techniques helps physicians to extract the features that may not be normally visible in a single image by different modalities. The usage of fusion in radiotherapy and skull surgery. Here, the information provided by magnetic resonance imaging (MRI) and X-ray computed tomography (CT) is complementary. CT provides best information about denser tissue and MRI offers best information on soft tissue. Normal and pathological soft tissues are better visualized by MRI, while the structure of tissue bone is better visualized by CT. The composite image, not only provides salient information from both images simultaneously, but also reveals the relative position of soft tissue with respect to the bone structure.

Fusion of images taken at different resolutions, intensity and by different techniques helps physicians to extract the features that may not be normally visible in a single image by different modalities. This work aims at the fusion of registered CT and MRI Images.

This fused image can significantly benefit medical diagnosis and also the further image processing such as, visualization (colorization), segmentation, classification and computer-aided diagnosis (CAD). Especially in image fusion the edge preservation is important in obtaining the complementary details of the input images. As an edge representation in Curvelet is better, Curvelet based image fusion is best suited for medical images. The images used here are gray scale CT and MRI images. However, the images of other modalities (like PET, SPECT, and X-ray etc.) with their true color nature may also be fused using the same methods. [1] In a wavelet-based image fusion algorithm the images to be fused are firstly decomposed into high frequency and low frequency bands. Then, the low frequency components are combined with the maximum energy rule and high frequency components are combined with variance rule. Finally, the fused image is constructed by inverse wavelet transform. They have selected four groups of images to simulate, and compare their simulation results with the pixel averaging method and most common wavelet method based on mean-max fusion rule. In the process of fusion they gave fusion rule based on energy and variance, which effectively conserved the energy of source images and avoided the loss of useful information. Comparing with the most common algorithm, the proposed fusion method gives improved visual effect of fused image and also improved objective parameter such as, entropy, standard deviation and RMSE.

[3] Techniques for image fusion like, primitive fusion (Averaging Method, Select Maximum, and Select Minimum), Discrete Wavelet transform based fusion, Principal component analysis (PCA) based fusion etc. are proposed. Comparison of all the techniques concludes the better approach for its future research. The Wavelet transforms is the very good technique for the image fusion provides a high quality spectral content. But a good fused image has both qualities so the combination of DWT & spatial domain fusion method (like PCA) fusion algorithm improves the performance as compared to use of individual DWT and PCA algorithm. Finally, this review concludes that image fusion algorithm based on a combination of DWT and PCA with morphological processing will improve the image fusion quality and may be the future trend of research regarding image fusion. [4].

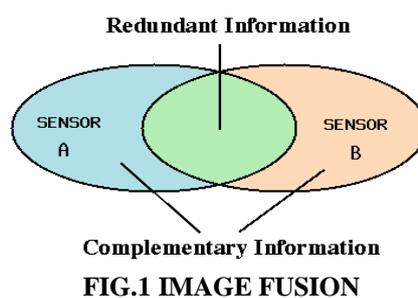
A novel wavelet based technique for image fusion which is developed by taking into account the physical meaning of the wavelet coefficients. After the images to be fused are decomposed by the wavelet transform, the coefficients of higher frequency bands and low frequency band are performed with new fusion schemes: the former is selected by a variety based scheme and the latter is selected by an edge based scheme. The performance of the proposed method is quantitatively compared with those of pixel averaging method and wavelet based methods. Experimental results clearly demonstrate the feasibility and effectiveness of the proposed method. [5].

The basic fusion algorithms, the pyramid based algorithms and the basic DWT algorithms are developed. The objective of the work was to assess the wide range of algorithms together, which is not found in the literature. The fused images were assessed using Structural Similarity Image Metric (SSIM), Laplacian Mean Squared Error along with seven other simple image quality metrics that helped us measure the various image features; Pareto Optimization method is used to figure out the algorithm that consistently had the image quality metrics produce the best readings Out of total eleven image fusion techniques were implemented, three very basic fusion techniques were Averaging Method, Maximum Selection Method and Minimum Selection Method, five pyramidal methods were FSD Pyramid, Laplacian Pyramid, Gradient Pyramid, Ratio Pyramid and Morphological Pyramid Methods and two of basic wavelet methods were Haar Wavelet and DBSS(2,2) Wavelet Methods. A set of nine image metrics were implemented to assess the fused image quality. The fused images of each set were also assessed based on their visual quality by ten respondents selected in random. The quality assessment based on the image metrics developed and visual perception was compared to assess the credibility of the image metrics. The readings produced by the 9 image metrics developed - MSE, PSNR, SC, NCC, AD, MD, NAE, LMSE and SSIM, were used to assess the best fusion algorithm (in terms of the quality of the fused images) using Pareto optimality method. DWT with Haar based fusion method was assessed best. [6].

Wavelet-based schemes perform better than standard schemes, particularly in terms of minimizing color distortion. Schemes that combine standard methods with wavelet transforms produce superior results than either standard methods or simple wavelet-based methods alone. The results from wavelet-based methods can also be improved by applying more sophisticated models for injecting detail information; however, these schemes often have greater set-up requirements. The simplest wavelet-based fusion scheme tends to produce better results than standard Fusion schemes such as IHS and PCA, further improvement is evident with more sophisticated wavelet based fusion schemes. The drawback is that there is greater computational complexity and often parameters must be set up before the fusion scheme can be applied.[7].

Authors have done the analysis of results of image fusion methods. As per their analysis fused image using pixel-based technique has good contrast and it works well for input images with similar levels of contrast. When input images have drastically different contrasts, pixel base fusion tends to take most data from the brighter image. Fused image has lower contrast in case of average based technique and it works well for input images with different levels of contrast. The wavelet based technique has very good contrast. In real applications, information from different sensors is not likely to be treated equally important. They have proposed a pixel based system that simultaneously compute image pixel intensity value taken from set of images and compared both and find the best value to be considered for the final fuse image. But problem of blurs is with this method and recovers the original undistorted image, all in high resolution, without any prior knowledge of the blurs and original image. Wavelet based method for image fusion gives better results which removes the limitation of pixel based image fusion method[8].

Image fusion based on wavelet decomposition, i.e. a multiresolution image fusion approach can fuse images with the same or different resolution level, i.e. range sensing, visual CCD, infrared, thermal or medical. A comparative analysis is carried out against classical existing strategies, including those of multiresolution. When the images are smooth, without abrupt intensity changes, the wavelets work appropriately, improving the results of the mentioned classical methods. This has been verified with smooth images and also with the medical images, where no significant changes are present. [9].



II. WAVELET BASED IMAGE FUSION

Wavelet theory is one of the most modern areas of mathematics. Wavelet transforms and other multiscale analysis functions have been used for compact signal and image representations in de-noising, compression and feature detection processing problems for about twenty years. Numerous research works have proven that space-frequency and space-scale expansions with this family of analysis functions provided a very efficient framework for signal or image data. The wavelet transform itself offers great design flexibility. Basis selection, spatial-frequency tiling, and various wavelet threshold strategies can be optimized for best adaptation to a processing application, data characteristics and feature of interest. Wavelet transformation, originally a mathematical tool for signal processing, is now popular in the field of image fusion. Recently, many image fusion methods based on wavelet transformation have been published. The wavelets used in image fusion can be categorized into three general classes: Orthogonal, Biorthogonal and Nonorthogonal. Although these wavelets share some common properties, each wavelet leads to unique image decomposition and a reconstruction method which leads to differences among wavelet fusion methods[10].

2.1. Wavelets for image fusion

1. It is a multiscale (multiresolution) approach well suited to manage the different image resolutions. Multiscale information can be useful in a number of image processing applications including the image fusion.
2. The discrete wavelets transform (DWT) allows the image decomposition in different kinds of coefficients preserving the image information.
3. Such coefficients coming from different images can be appropriately combined to obtain new coefficients, so that the information in the original images is collected appropriately.
4. Once the coefficients are merged, then the fused image is achieved through the inverse discrete wavelets transform (IDWT), where the information in the merged coefficients is also preserved. The key step in image fusion based on wavelets is that of coefficient combination, namely, the process of merge the coefficients in an appropriate way in order to obtain the best quality in the fused image. This can be achieved by a set of strategies. The most simple is to take the average of the coefficients to be merged, but there are other merging strategies with better performances. [16]

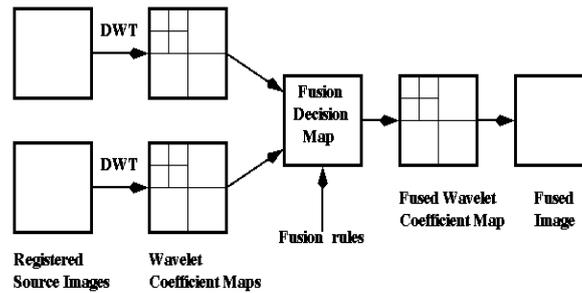


Fig.1 Fusion process using multi level image decomposition.

2.2. Discrete wavelet transform (DWT)

Fusion based on transforms has some advantages over other simple methods, like: energy compaction, Larger SNR, Reduced features, etc. The transform coefficients are representative for image pixels. Wavelets are used for time frequency localization, and perform multi-scale and multi resolution operations. Discrete wavelet transform (DWT), transforms a discrete time signal to a discrete wavelet representation. It converts an input series x_0, x_1, \dots, x_m , into one high-pass wavelet coefficient series and one low-pass wavelet coefficient series of length $n/2$ each given by

$$H_i = \sum_{m=0}^{k-1} x_{2i-m} * S_m(Z) \quad \dots\dots\dots (1)$$

$$L_i = \sum_{m=0}^{k-1} x_{2i-m} * t_m(Z) \quad \dots\dots\dots (2)$$

$sm(z)$ and $tm(z)$ are called wavelet filters, k is the length of the filter, and $i=0, \dots, [n/2]-1$. In practice, such a transformation will be applied recursively on the low-pass series until the desired number of iterations is reached.

2.3. The advantages of haar wavelet transform as follows

Best performance in terms of computation time.

1. Computation speed is high.
2. Simplicity
3. HWT is efficient compression method
4. It is memory efficient, since it can be calculated in place without a temporary array

III. IMAGE FUSION RULE

Fusion rules determine how the source transforms will be combined:

- Fusion rules may be application dependent
 - Fusion rules can be the same for all sub-bands or dependent on which sub-band is being fused
- There are two basic steps to determine the rules:
- compute salience measures corresponding to the individual source Transforms
 - decide how to combine the coefficients after comparing the salience measures (selection or averaging)
 - The main issue in wavelet fusion technique is the selection of decomposition

Different modalities can be fused using the Fusion rule. Fusion selection rule includes choosing the salient features of the image inputs. Higher absolute values of coefficients correspond to features such as edges or singularities.

OBJECTIVES OF THE RESEARCH WORK

1. Using haar wavelet filter to preserve the edges and details of the images
2. To efficiently reduce noise in medical images,
3. Comparison of various wavelet based pixel image fusion method to achieve minimum Mean Square Error value and maximum Signal to Noise Ratio value
4. Results of various fusion techniques are extracted and then these results are again compared with other Image Fusion methods and find the best suited method for medical applications.

IV. PIXEL LEVEL FUSION

Pixel-by-pixel fusion techniques include the basic arithmetic operations, logic operations and probabilistic operations as well as slightly more complicated mathematical operations. The image values include pixel gray-levels, feature map values and decision map labels. Although more sophisticated techniques are available, the simple pixel operations are still widely used in many image fusion applications. We consider fusion techniques which rely on simple pixel operations on the input image values. We assume the input images are spatially and temporally aligned, semantically equivalent and radiometrically calibrated. The image fusion of K input images I_1, I_2, \dots, I_K using a simple arithmetic addition operator.

4.1. LINEAR FUSION RULE: ADDITION

Addition which is probably the simplest fusion operation. It works by estimating the average intensity value of the input images $I_k, \{k=1, 2, \dots, K\}$, on a pixel-by-pixel basis. If $I(m, n)$ denotes the fused image at the pixel (m, n) , then

$$I(m, n) = \frac{1}{K} \sum_{k=1}^K I_k(m, n) \quad \dots\dots\dots(1)$$

4.2. Nonlinear Fusion Rule: Nonlinear Methods

Another simple approach to image fusion is to build the fused image by the application of a simple nonlinear operator such as max or min. If in all input images the bright objects are of interest, a good choice is to compute the fused image by a pixel-by-pixel application of the maximum operator.

4.2.1. Maximum fusion rule:

$$I(m, n) = \max(I_1, I_2, \dots, I_k); \quad \dots\dots\dots(2)$$

4.2.2. Minimum fusion rule:

$$I(m, n) = \min(I_1, I_2, \dots, I_k); \quad \dots\dots\dots(3)$$

Image fusion aims at the integration of various complementary image data into a single, new image with the best possible quality. The term "quality" depends on the demands of the specific application, which is usually related to its usefulness for human visual perception, computer vision or further processing.

V. OBJECTIVE EVALUATION USING STATISTICAL CHARACTERISTICS

Some of the statistical measures are Signal to Noise Ratio (SNR), Peak signal to noise ratio (PSNR) and Mean square error (MSE). These are commonly used measures in assessing image fusion techniques that consider an image as a special type of signal. The quality of signal is often expressed quantitatively with the signal-to-noise ratio defined as

$$SNR = 10 \log_{10} \frac{\sum_{m=1}^{S_1} \sum_{n=1}^{S_2} z(m, n)^2}{\sum_{m=1}^{S_1} \sum_{n=1}^{S_2} [z(m, n) - o(m, n)]^2} \quad \dots\dots\dots(4)$$

Where $z(m, n)$ and $o(m, n)$ denote the intensity of the pixel of the estimated and original image respectively at location (m, n) . The size of the image is $S_1 \times S_2$. High values of SNR show that the error of the estimation is small and therefore, among various image fusion methods the ones that exhibits higher SNR's can be considered of better performance. The PSNR and MSE are measures similar to the SNR defined as

$$PSNR = 10 \log_{10} \frac{255^2}{\sum_{m=1}^{S_1} \sum_{n=1}^{S_2} [z(m, n) - o(m, n)]^2} \quad \dots\dots\dots(5)$$

$$MSE = \frac{\sum_{m=1}^{S_1} \sum_{n=1}^{S_2} [z(m, n) - o(m, n)]^2}{255^2} \quad \dots\dots\dots(6)$$

When assessing the performance of an image fusion technique using the above mentioned measurements, we require knowledge of the original image (ground truth or full-referenced). For the reason these measurements can be used only with synthetic (simulated) data. The above measurements exhibit the drawback of providing a global idea regarding the quality of an image. In cases where the fused image exhibits artifacts concentrated within a small area, these measurements can still provide an acceptable value even if the image is visually unacceptable.

VI. OBJECTIVE EVALUATION BASED ON IMPORTANT FEATURES

One goal of image fusion is to integrate complementary information from multiple sources so that the new images are more suitable for the purpose of human visual perception and computer processing. Therefore, a measure should estimate how much information is obtained from the input images.

Different quality measures

1. LABF: Fusion loss
2. NABF: fusion artifacts or fusion noise
3. QABFc common information in fused image.
4. QdABF: total fusion gain.
5. Qc common information component compute the
6. QABF :information contribution of
7. QAFo information contribution of A into F
8. QBFo information contribution of B into F
9. Qd = abs(QAF - QBF);
10. Qc = (QAF+QBF-Qd)/2 : common information component
11. QdAF = QAF - Qc; QdBF = QBF - Qc;

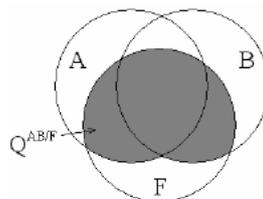


Figure 6: Graphical representation of the image information fusion process

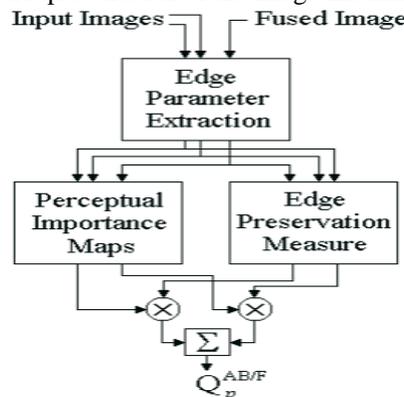


Figure 7: Basic structure of the objective image fusion performance measure

5. QUALITY METRICS IMPLEMENTED

- a) Statistical measures: SNR, PSNR, MSE
- b) Edge based quality measures: $Q^{AB/F}$, $Q_c^{AB/F}$, $Q_d^{AB/F}$, and $L^{AB/F}$
- c) UIQI based quality measure: Q_o

For UIQI based quality measures the following specifications have been taken:

- 1) Window of all ones of size 8X8 is taken.
- 2) Contrast is taken as the saliency for the input images. Variance corresponds to the contrast. So variance is calculated for every sliding window.

$$Q_o = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2xy}{x^2 + y^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \dots\dots\dots(7)$$

Wang and Bovik refer to Q_o as an *image quality index* and use it to quantify the structural distortion between images x and y , one of them being the reference image and the other the distorted one. In fact, the value $Q_o = Q_o(x, y)$ is a measure for the similarity of images x and y and takes values between -1 and 1. Note that the first component in eqn.(7) is the correlation coefficient between x and y . The second component corresponds to a kind of average luminance distortion and it has a dynamic range of [0, 1] (assuming nonnegative mean values). The third factor in eqn.(7) measures a contrast distortion and its range is also [0, 1]. The maximum value $Q_o = 1$ is achieved when x and y are identical.

Common information in F (fused image):

$$Q_c^{AB/F} = \frac{\sum_{r,n,m} Q_{r,n,m}^c (w_{r,n,m}^A + w_{r,n,m}^B)}{\sum_{r,n,m} [w_{r,n,m}^A + w_{r,n,m}^B]} \dots\dots\dots(8)$$

Fusion loss

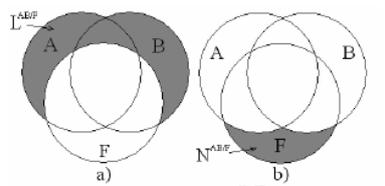


Figure 8: a) Fusion information loss $L^{AB/F}$ and b) Fusion artifacts, $N^{AB/F}$

RESULTS BASED ON MEDICAL IMAGE FUSION

Table 9.1

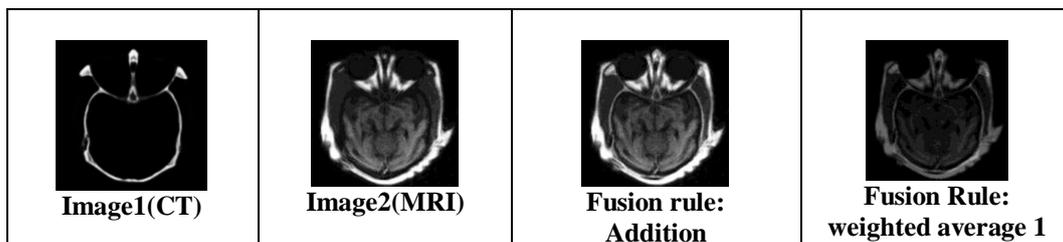
FUSION RULE	Mean	SNR	MSE	PSNR (db)	RMSE	Std.Dev.
Addition	59.62	0.738	6142.88	10.281	78.37	60.58
weighted avg1 (0.6 *CT + 0.4*MRI)	25.07	1.476	954.975	18.364	30.90	26.50
weighted avg2 (0.4 *CT + 0.6*MRI)	34.84	0.661	2152.30	14.835	46.39	35.68
Subtraction (max pix -min. pixel)	52.57	0.120	5808.20	10.524	76.21	56.22
Pixel Min	4.66	1.809	139.48	26.719	11.81	13.71
Pixel Max	56.18	0.475	5815.22	10.519	76.25	57.27

Table 9.2

FUSION RULE	QABF	QABFC	QdABF	NAB	Quality index Q0	LABF
Addition	0.7941	0.0206	0.773	0	0.0696	0.131
weighted avg1	0.2567	0.0448	0.211	0	0.1497	0.738
weighted avg2	0.5698	0.0480	0.521	0	0.1182	0.422
Subtraction	0.5850	0.02662	0.558	0	0.1089	0.294
Min	0.1266	0.0455	0.081	0	0.4401	0.855
Max	0.7461	0.0488	0.697	0	0.0945	0.199

Table 9.3

FUSION RULE	Joint histogram_medA	Similarity_medA	RMSE_wrt_medA	Correlation_medA	Mutual information_medA	Similarity_medB
Add	8.031	0.089	14.80	0.363	0.442	0.899
wtd avg1	6.905	0.181	11.66	0.506	0.452	0.550
wtd avg2	7.464	0.106	12.77	0.277	0.373	0.800
Sub	8.056	0.032	14.07	0.006	0.322	0.827
Min	3.891	0.799	2.88	0.835	0.604	0.064
Max	8.029	0.082	14.32	0.204	0.377	0.909



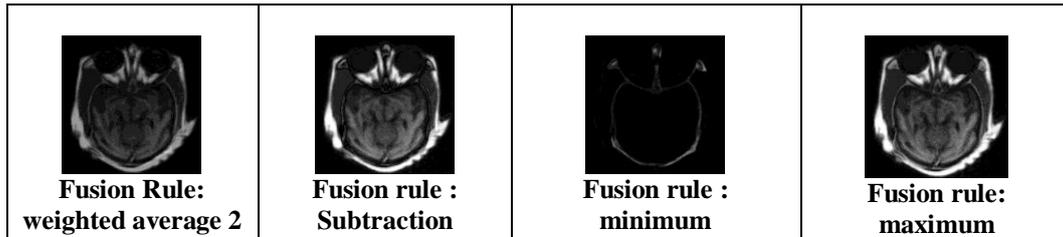
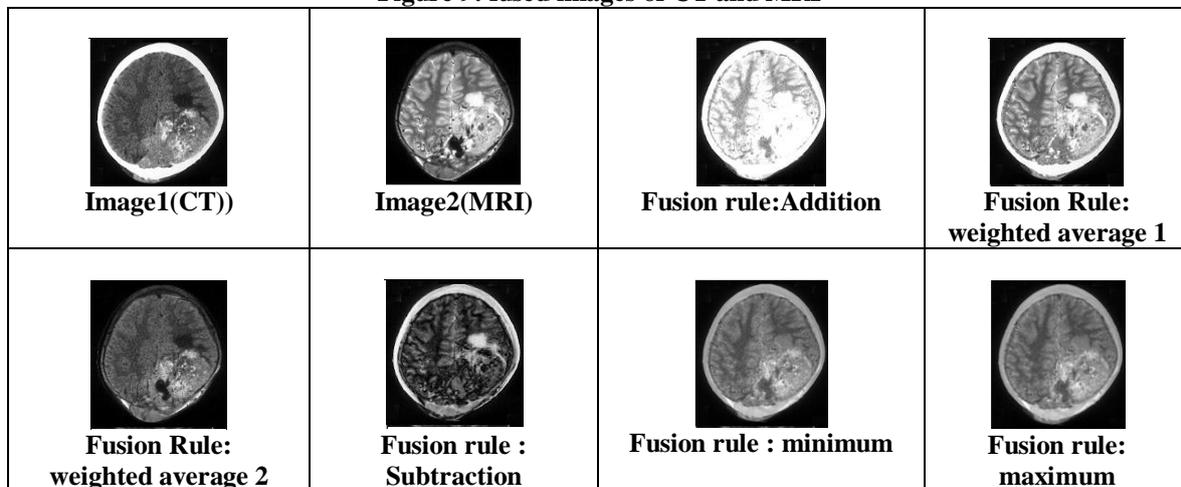


Figure 9: fused images of CT and MRI



VII. DISCUSSIONS AND CONCLUSION

The wavelet based fusion method using different fusion rules is proposed and it is applied on the CT and MRI images of the human brain. We have used different sets of the registered multimodal brain images. It is observed that minimum pixel replacement method is giving the better results as compared to other methods as the noise is much reduced in minimum method. Contrast of the images is much improved in the coefficient addition method. Also if we change the weights on the both modality we are getting different results. Contrast improves if we are giving more weight to MRI image (i.e. soft tissue structure visible more clearly). Error observed is less when more weight is given to CT. Any artifacts present in the fusion process may be considered as information. As medical images are low contrast images the information in all the parts (intensity areas) is equally important. We have also assess the fused images using edge based quality parameters also to observe how much edge information is preserved in the fused image. Edge is more efficiently preserved in the coefficient addition method of fusion with less information loss. The information retrieval is of much importance in medical application. The information in the soft tissue and bony structure both are important in diagnosis of any canceral development or any fracture. Contrast of the images.

VIII. FUTURE SCOPE

In many important imaging applications, images exhibit edges and discontinuities across curves. In biological imagery, this occurs whenever two organs or tissue structures meet. Especially in image fusion the edge preservation is important in obtaining the complementary details of the input images. As edge representation in Curvelet is better as it enhancing the curved edges more efficiently whereas in wavelet it can enhance the linear edges better than curved edges, Curvelet based image fusion is best suited for medical images. [10]

REFERENCES

- [1] KiranParmar, Rahul Kher "A Comparative Analysis of Multimodality Medical Image Fusion Methods"- 2012 Sixth Asia Modelling Symposium 978-0-7695-4730-5/12 2012 IEEE PP 93-97 DOI 10.1109/AMS.2012.46
- [2] Kiran Parmar,Asso. Prof. Rahul K Kher, Asst. Prof. Falgun N Thakkar-"Analysis of CT and MRI Image Fusion using Wavelet Transform"- 2012 International Conference on Communication Systems and Network Technologies
- [3] Mrs.S.V.More, Prof.Dr.Mrs.S.D.Apte -" Pixel-Level Image Fusion Using Wavelet Transform"- International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 5, July - 2012 ISSN: 2278-0181
- [4] Deepak Kumar Sahu, M.P.Parsai-" Different Image Fusion Techniques –A Critical Review"- International Journal of Modern Engineering Research (IJMER) Vol. 2, Issue. 5, Sep.-Oct. 2012 pp-4298-4301 (www.ijmer.com)

- [5] Yong Yang-"Multiresolution Image Fusion Based on Wavelet Transform By Using a Novel Technique for Selection Coefficients" -JOURNAL OF MULTIMEDIA, VOL. 6, NO. 1, FEBRUARY 2011
- [6] Shivsubramani Krishnamoorthy, K P Soman -"Implementation and Comparative Study of Image Fusion Algorithms "- International Journal of Computer Applications (0975 – 8887) Volume 9– No.2, November 2010
- [7] Krista Amolins, Yun Zhang, Peter Dare- "Wavelet based image fusion techniques — An introduction, review and comparison"- ISPRS Journal of Photogrammetry & Remote Sensing 62 (accepted 19 May 2007) 249–263 www.elsevier.com/locate/isprsjprs
- [8] Chetan K. Solanki and Narendra M. Patel-" Pixel based and Wavelet based Image fusion Methods with their Comparative Study"- National Conference on Recent Trends in Engineering & Technology
- [9] Gonzalo Pajares, Jesus Manuel de la Cruz" A wavelet-based image fusion tutorial" -Pattern Recognition 37 (2004) 1855 – 1872 www.elsevier.com/locate/patcog
- [10] M. Deshmukh, U. Bhosale, "Image Fusion and Image Quality Assessment of Fused Images", International Journal of Image Processing (IJIP), Volume (4): Issue (5) 484-508, 2009
- [11] M.A.Mohamed, A. S. Asem, "Medical image filtering, fusion and classification techniques", Egyptian Journal of Bronchology, Vol. 5, No 2, December, 2011 Dec 2011.
- [12] R. Singh, M. Vatsa, A. Noore, "Multimodal Medical Image Fusion using Redundant Discrete Wavelet Transform", 2009.
- [13] L. Chiorean, M. Vaida, "Medical Image Fusion Based on Discrete Wavelet Transform Using Java Technology" ITI 2009.
- [14] Xydeas, C., and Petrovic, V., "Objective Pixel-level Image Fusion Performance Measure," Sensor Fusion: Architectures, Algorithms, and Applications IV, SPIE 4051:89-98, 2000.
- [15] H. Li, B. S. Manjunath and S. K. Mitra "Multisensor Image Fusion Using the Wavelet Transform", Graphical Models and Image Processing, 57 (3): 235-245, 1995.
- [16] Lau Wai Leung, Bruce King and Vijay Vohora, "Comparison of Image Fusion Techniques uses Entropy and INI", In: Pro. 22nd Asian Conference on Remote Sensing, 5-9 Nov 2001.
- [17] Xydeas, C., and Petrovic, V., "Objective Pixel-level Image Fusion Performance Measure," Sensor Fusion: Architectures, Algorithms, and Applications IV, SPIE 4051:89-98, 2000.
- [18] Y. Kiran Kumar, "Comparison Of Fusion Techniques Applied To Preclinical Images: Fast Discrete Curvelet Transform Using Wrapping Technique & Wavelet Transform" Journal Of Theoretical And Applied Information Technology, 2003.
- [19] L.W.Leung, B.King, V. Vohora, "Comparison of image data fusion techniques using entropy and INI", CRISP, 2001.
- [20] Chu-Hui Lee and Zheng-Wei Zhou, "Comparison of Image Fusion based on DCT-STD and DWT-STD" proceedings of International Multiconference of engineers and Computer scientists 2012 Vol I, IMECS 2012 march 14-16 2012, Hong Kong
- [21] Gonzalo Pajares, Jesus Manuel de la Cruz, "A wavelet-based image fusion tutorial", Pattern Recognition 37 (2004) 1855 – 1872
- [22] Deepak Kumar Sahu, M.P.Parsai-"Different Image Fusion Techniques –A Critical Review"- International Journal of Modern Engineering Research (IJMER) Vol. 2, Issue. 5, Sep.-Oct. 2012 pp-4298-4301
- [23] <http://www.bic/mni.mcgill.ca/brainweb—medical images>
- [24] <http://www.imagefusion.org>

Author's Biography:



Ms. Deepali Sale is pursuing her PhD in image processing (image fusion). She has completed her M.E. (Electronics) from College of Engineering, Pune. She is presently working as an Assistant Professor in Pad. Dr. D.Y. Patil Institute of Engineering and Technology, Pimpri, Pune. She has about 10 papers published in national/ international Journals and conferences. She is a member of IEEE and IETE, India



Dr. M.A. Joshi has completed her Ph.D. from Pune University and Msc. (Tech) by research from the University of Manchester, U.K., M.E and B.E. (E & TC) from Pune University. She is currently working as a Professor at College of Engineering, Pune (COEP) an autonomous. Institute of Government of Maharashtra. She has a teaching experience of about 30 years and Industry experience of about 1 year. She has about 105 papers published in national / international conferences and journals. She is a fellow of THEIET and member of IEEE.



Ms. Varsha Patil is pursuing her M.E. Electronics & Telecommunication in communication network from Pad. Dr. D.Y. Patil Institute of Engineering and Technology, Pimpri, Pune. She has about one year teaching experience in Pad. Dr. V. B. Kolte college of engineering, Malkapur.



Ms. Pallavi Sonare is pursuing her M.E. Electronics & Telecommunication in communication network from Pad. Dr. D.Y. Patil Institute of Engineering and Technology, Pimpri, Pune.



Mrs. Chaya Jadhav completed her M.E. (Electronics) from College of Engineering, Pune. Presently she is working as an assistant professor in Pad. Dr. D.Y. Patil Institute of Engineering and Technology, Pimpri, Pune. She has about 12 years of teaching experience

Content Based Image Retrieval in Medical Imaging

Prachi.Bhende¹, Prof.A.N.Cheeran²

¹Department of Electrical Engineering, VJTI, Mumbai, India

²Department of Electrical Engineering, VJTI, Mumbai, India

ABSTRACT

The advancement in the field of medical imaging system has lead industries to conceptualize a complete automated system for the medical procedures, diagnosis, treatment and prediction. The success of such system largely depends upon the robustness, accuracy and speed of the retrieval systems. Content based image retrieval (CBIR) system is valuable in medical systems as it provides retrieval of the images from the large dataset based on similarities. There is a continuous research going on in the area of CBIR systems typically for medical images, which provides a successive algorithm development for achieving generalized methodologies, which could be widely used. The aim of this paper is to discuss the various techniques, the assumptions and its scope suggested by various researchers and further setup a roadmap for research in the field of CBIR system for medical image database.

INDEX TERMS: CBIR, Medical Imaging

I. INTRODUCTION

Content-based image retrieval (CBIR) is the application of computer vision techniques to the problem of digital image search in large databases. CBIR enables to retrieve the images from the databases [1, 2]. Medical images are usually fused, subject to high inconsistency and composed of different minor structures. So there is a necessity for feature extraction and classification of images for easy and efficient retrieval [3]. CBIR is an automatic retrieval of images generally based on some particular properties such as color, Composition, shape and texture [4, 5]. Every day large volumes of different types of medical images such as dental, endoscopy, skull, MRI, ultrasound, radiology are produced in various hospitals as well as in various medical centres [6]. Medical image retrieval has many significant applications especially in medical diagnosis, education and research fields. Medical image retrieval for diagnostic purposes is important because the historical images of different patients in medical centres have valuable information for the upcoming diagnosis with a system which retrieves similar cases,

Make more accurate diagnosis and decide on appropriate treatment. The term content based image retrieval was introduced by Kato[7], while describing his experiment of image retrieval from a database on the basis of color and shape features. There is a significant amount of growing image databases in medical field images. It is a proven though that for supporting clinical decision making the integration of content based access method into Picture Archiving and Communication Systems (PACS) will be a mandatory need [8]. In most biomedical disciplines, digital image data is rapidly expanding in quantity and heterogeneity, and there is an increasing trend towards the formation of archives adequate to support diagnostics and preventive medicine. Exploration, exploitation, and consolidation of the immense image collections require tools to access structurally different data for research, diagnostics and teaching. Currently, image data is linked to textual descriptions, and data access is provided only via these textual additives. There are virtually no tools available to access medical images directly by their content or to cope with their structural differences. Therefore, visual based (i.e. content-based) indexing and retrieval based on information contained in the pixel data of biomedical images is expected to have a great impact on biomedical image databases. However, existing systems for content-based image retrieval (CBIR) are not applicable to the biomedical imagery special needs, and novel methodologies are urgently needed.

Content based image retrieval (CBIR) has received significant attention in the literature as a promising technique to facilitate improved image management in PACS system [9, 10]. The Image Retrieval for Medical Applications (IRMA) project [10,11] aims to provide visually rich image management through CBIR techniques applied to medical images using intensity distribution and texture measures taken globally over the entire image.

This approach permits queries on a heterogeneous image collection and helps in identifying images that are similar with respect to global features. Section 2 highlights about the significance of CBIR in medical imaging followed by methods used for implementation CBIR in Section 3. The recent work done on CBIR is mentioned in Section 4. The issues or research gap from prior work is illustrated in Section 5 followed conclusion in Section 6.

II. SIGNIFICANCE OF CBIR IN MEDICAL IMAGING

There are several reasons why there is a need for additional, alternative image retrieval methods apart from the steadily growing rate of image acquired every day. It is important to explain these needs and to discuss possible technical and methodological improvements and the resulting clinical benefits. The goals of medical information systems have often been defined to deliver the needed information at the right time, the right place to the right persons in order to improve the quality and efficiency of care processes [12]. Such a goal will most likely need more than a query by patient name, series ID or study ID for images. For the clinical decision making process it can be beneficial or even important to find other images of the same modality, the same anatomic region of the same disease. Although part of this information is normally contained in the Digital Imaging and Communication in Medicine (DICOM) headers and many imaging devices are DICOM compliant at this time, there are still some problems. DICOM headers have proven to contain a fairly high rate of errors, for example for the field anatomical region, error rates of 16% have been reported [13]. This can hinder the correct retrieval of all wanted images. Clinical decision support techniques such as case based reasoning [14] or evidence based medicine [15,16] can even produce a stronger need to retrieve images that can be valuable for supporting certain diagnoses. It could even be imagined to have Image Based Reasoning (IBR) as a new discipline for diagnostic aid. Decision support systems in radiology [17] and computer aided diagnostics for radiological practice as demonstrated at the RSNA (Radiological Society of North America) [18] are on the rise and create a need for powerful data and metadata management and retrieval.

The general clinical benefit of imaging system has also already been demonstrated by B. Kaplan et.al [19]. An initiative is described by A. Horsch et.al to identify important tasks for medical imaging based on their possible clinical benefits. It needs to be stated that the purely visual image queries as they are executed in the computer vision domain will most likely not be able to ever replace text based methods as there will always be queries for all images of a certain patient, but they have the potential to be a very good complement to text based search based on their characteristics. Still, the problems and advantages of the technology have to be stressed to obtain acceptance and use of visual and text based access methods up to their full potential. A scenario for hybrid, textual and visual queries is put forward for the CBIR system by S. Antani et.al [21]. Besides diagnostics, teaching and research especially are expected to improve through the use of visual access methods as visually interesting images can be chosen and can actually be found in the existing large repositories. The inclusion of visual features into medical studies is another interesting point for several medical research domains. Visual features do not only allow the retrieval of cases with patients having similar diagnoses but also cases with visual similarity but different diagnoses. In teaching it can help lecturers as well as students to browse educational image repositories and visually inspect the results found. This can be the case for navigating in image atlases. It can also be used to cross correlate visual and textual features of the images.

III. METHODS USED FOR IMPLEMENTING CBIR

Content-based image retrieval hinges on the ability of the algorithms to extract pertinent image features and organize them in a way that represents the image content. Additionally, the algorithms should be able to quantify the similarity between the query visual and the database candidate for the image content as perceived by the viewer. Thus, there is a systemic component to CBIR and a more challenging semantic component.

- □ Shape Based Method: For shape based image retrieval, the image feature extracted is usually an N dimensional feature vector which can be regarded as a point in a N dimensional space. Once images are indexed into the database using the extracted feature vectors, the retrieval of images is essentially the determination of similarity between the query image and the target images in database, which is essentially the determination of distance between the feature vectors representing the images. The desirable distance measure should reflect human perception. Various similarity measures have been exploited in image retrieval. In one of the implementation by Muller .H they have used Euclidean distance for similarity measurement.[22]
- □ Texture Based Method: Texture measures have an even larger variety than color measures. Some common measures for capturing the texture of images are wavelets and Gabor filters. The texture measures try to capture the characteristics of images or image parts with respect to changes in certain directions and scale of changes. This is most useful for regions or images with homogeneous texture.

□ □ Continuous Feature Selection Method: This method deals with the “dimensionality curse” and the semantic gap problem. The method applies statistical association rule mining to relate low-level features with high-level specialist’s knowledge about the image, in order to reduce the semantic gap existing between the image representation and interpretation. These rules are employed to weigh the features according to their relevance. The dimensionality reduction is performed by discarding the irrelevant features (the ones whose weight are null). The obtained weights are used to calculate the similarity between the images during the content-based searching. Experiments performed show that the proposed method improves the query precision up to 38%. Moreover, the method is efficient, since the complexity of the query processing decreases along the dimensionality reduction of the feature vector. □ □ With Automatically Extracted MeSH Terms: There is still a semantic gap between the low-level visual features(textures,colors) automatically extracted and the high level concepts that users normally search for (tumors, abnormal tissue)[22].

Some solutions stated by Juan C Caicedo et.al to bridge the semantic gap are the connection of visual features to known textual labels of the images [23] or the training of a classifier based on known class labels and the use of the classifier on unknown cases is also discussed by Jia Li et.al.[24].Combinations of textual and visual features for medical image retrieval have as of yet rarely been applied, although medical images in the electronic patient record or case databases basically always do have text attached to them. The complementary nature of text and visual image features for retrieval promises to lead to good retrieval results. □ □ Using Low Level Visual Features and The image retrieval process consists of two main phases: pre-processing phase and retrieval phase. Both phases are described as follows. The pre-processing phase is composed of two main components: a feature extraction model and a classification model. The input of the pre-processing phase is the original image database, i.e. images from the ImageCLEFmed collection, with more than 66,000 medical images. The output of the pre-processing phase is an index relating each image to its modality and a feature database. □ □ The Feature Extraction Model: The feature extraction model operates on the image database to produce two kinds of features: histogram features and Meta features. Histogram features are used to build the feature database, which is used in the retrieval phase to rank similar images. Met features are a set of histogram descriptors, which are used as the input to the classification model to be described later. Histogram features used in this system are:

- o Gray scale and color histogram (Gray and RGB)
- o Local Binary Partition histogram (LBP)
- o Tamura texture histogram (Tamura)
- o Sobel histogram (Sobel)
- o Invariant feature histogram (Invariant) Meta features are calculated from histogram features in order to reduce the dimensionality. These meta features are the four moments of the moment generating function (mean, deviation, skewness and kurtosis) and the entropy of the histogram. Each histogram has five associated meta features, meaning a total of 30 meta features with information of color, texture, edges and invariants.

IV. RECENT WORK IN CBIR

Support vector machines (SVM) are extensively used to learn from relevance feedback due to their capability of effectively tackling the above difficulties. However, the performances of SVM depend on the tuning of a number of parameters. It is a different approach based on the nearest neighbour paradigm. Each image is ranked according to a relevance score depending on nearest neighbour distances. This approach allows recalling a higher percentage of images with respect to SVM-based techniques [25] there after quotient space granularity computing theory into image retrieval field, clarify the granularity thinking in image retrieval, and a novel image retrieval method is imported. Firstly, aiming at the Different behaviours under different granularities, obtain color features under different granularities, achieve different quotient spaces; secondly, do the attribute combination to the obtained quotient spaces according to the quotient space granularity combination principle; and then realize image retrieval using the combined attribute function.[26] Then a combination of three feature extraction methods namely color, texture, and edge histogram descriptor is reviewed. There is a provision to add new features in future for better retrieval efficiency. Any combination of these methods, which is more appropriate for the application, can be used for retrieval. This is provided through User Interface (UI) in the form of relevance feedback. The image properties analyzed in this work are by using computer vision and image processing algorithms.

1. Evaluating an emotional response to color images. It is mainly used for the case – base reasoning methodology, emotional evolution of color images values , and also find out fuzzy similarity relational & inter and intra similarities and used for MPEG -7 visual descriptors. [27]

2. 3D Object: The 3D objects make their efficient retrieval technology highly desired. Intelligent query methodology, multiple view and representative query view. [28]
3. Relevance Feedback: Another methodology is classify the query in text or images to relevance / irrelevance set of images to select the positive images. Reference to retrieve the relevance images from databases. [29]

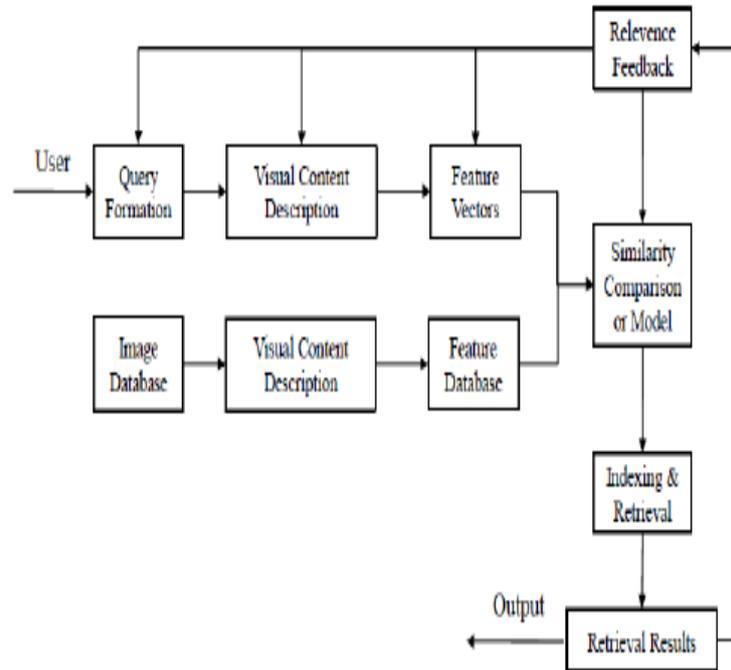


Figure: Block diagram of CBIR system

V. RESEARCH GAP

There are various areas to work with for the improvement of the content based image retrieval system. It is already been discussed that the existing techniques may be used to improve the quality of image retrieval and the understanding of user intentions. An approach that combines two different approaches to image retrieval, together with active use of context information and interaction has been proposed. The problem of bridging the semantic gap between high level query which is normally in terms of an example image and low level features of an image such as color, texture, shape and object forced to apply techniques to reduce the semantic gap. One approach to making a fair assessment of the state of the field is by comparing CBIR applications presented in the literature.

Survey on recent researches on implementation of CBIR in Medical imaging to the nature and content of the data, highlights that it is necessary to develop comparison methods that analyze more than the selection of particular techniques and the experimental results presented in the literature. Rather, it may be better to formally describe an idealized CBIR system and identify the shortcomings in the candidate system. These shortcomings have been labelled as “gaps” and extensively discussed by L.R.Long et.al in [30]. The concept of the gap is a generalization of the well-known “semantic gap” that refers to the difficulty of capturing high-level imaged content semantics from extracted low-level image features. These gaps have been broadly categorized into four types and defined below:

1. The Content Gap addresses a system’s ability to foster human understanding of concepts from extracted features. In medical applications, it also refers to the extent to which the system adapts to varying modalities, context, and diagnostic protocols.
2. The Feature Gap addresses the extent to which the image features are extracted. This is measured along several dimensions: degree of automation, degree of detail captured along the content axis (object structure), use of multi-scalar techniques, the use of space and (if available) time dimension in image data, and use of all channels on each dimension.

3. The Performance Gap : It addresses practicalities of system implementation and acceptance. It evaluates system availability, extent of integration into the medical infrastructure, use of feature indexing techniques, and the extent to which the system was evaluated.
4. The Usability Gap: It measures the richness of available query features and the extent to which they can be combined, available support for comprehending the results returned by the system, and available support for query refinement.

Addressing these aspects makes a CBIR system more usable, and may increase its acceptance into the medical (clinical, research, or education) workflow.

VI. CONCLUSION

This survey is highlighting the significant contributions of content based image & information's Retrieval field. The difficulty faced by CBIR methods in making in-roads into medical applications can be attributed to a combination of several factors. Some of the leading causes can be categorized according to the "gaps" model presented above. An idealized system can be designed to overcome all the above gaps, but still fall short of being accepted into the medical community for lack of (i) useful and clear querying capability; (ii) meaningful and easily understandable responses; and (iii) provision to adapt to user feedback. The opposite is also true to some extent. A technically mediocre, but promising, system may obtain valuable end user feedback, and by technical improvement may increase user acceptance with the application of usability design principles. Other than item (iii), which still needs significant research effort, the usability gap can only be bridged by keeping the end user in mind from early system development, as well as by conducting well designed usability studies with targeted users. In general, a high involvement of the user community in system design and development can significantly improve adoption and acceptance. The preceding subsections already showed the large variability in techniques that are used for the retrieval of images. Still, several very successful techniques from the image retrieval domain have not been used for medical images as of yet. The entire discussion on relevance feedback that first improved the performance of text retrieval systems and then, 30 years later, of image retrieval systems has not at all been discussed for the medical domain.

REFERENCES

- [1]. M.Smeulders, Worring, and M. Santini, "Content-based image Retrieval at The End of Early Years", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 22, No.12, 2000, pp. 1349-1380.
- [2]. V.S. Murthy, E.Vamsidhar, J.N.V.R. Swarup Kumar, and P. Sankara Rao, "Content based Image Retrieval using Hierarchical and K means Clustering Techniques", International Journal of Engineering Science and Technology, Vol. 2, No. 3, 2010, pp. 209-212.
- [3]. B. Ramamurthy, and K.R. Chandran, "CBMIR:Shape-based Image Retrieval using Canny Edge Detection and K-means Clustering Algorithms for Medical Images", International Journal of Engineering Science and Technology, Vol. 3, No. 3, 2011, pp. 209-212.
- [4]. Roberto Parades, Daniel Keysers, Thomas M. Lehman, Berthold Wein, Herman Ney, and Enrique Vidal, "Classification of Medical Images Using Local Representation", Workshop Bildverarbeitung für die Medizin, 2002, pp.171-174.
- [5]. Wei Zhang, Sven Dickinson, Stanley Sclaroff, Jacob Feldman, and Stanley Dunn, "Shape -Based Indexing in a Medical Image Database", Biomedical Image Analysis, 1998, pp. 221- 230.
- [6]. Monireh Esnaashari, S. Amirhassan Monadjami, and Gholamali Naderian, "A Content-based Retinal Image Retrieval Method for Diabetes- Related Eye Diseases Diagnosis", International Journal of Research and Reviews in Computer Science(IJRRCS), Vol. 2, No. 6, 2011, pp. 1222-1227.
- [8]. Kato, T., "Database architecture for content based image retrieval in Image Storage and Retrieval Systems" (Jambardino A and Niblack W eds), Proc SPIE 2185, pp 112-123, 1992.
- [9]. David Bandon, Christian Lovis, Antoine Geissbuhler, Jean-Paul Vallée, Enterprise-wide PACS: Beyond Radiology, an Architecture to Manage All Medical Images, AUR, 2005 doi:10.1016/j.acra.2005.03.075
- [10]. H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications-Clinical benefits and future directions", International Journal of Medical Informatics, Vol. 73, No. 1, 2004, pp. 1-23.
- [11]. T.M. Lehmann, M.O. Guld, C Thies, B Fischer, K. Spitzer, and D. Keysers, "Content-based image retrieval in medical applications", Methods of Info in Med, IOS Press , Vol. 43, No. 4, 2004, pp. 354-361.
- [12]. C. Thies, M.O. Guld, B Fischer, and T.M. Lehmann, "Content-based queries on the CasImage database within the IRMA framework", Lecture Notes in Computer Science, Springer 3491, 2005, pp. 781-792.
- [13]. A. Winter, R. Haux, A three level graph-based model for the management of hospital information systems, Methods of Information in Medicine 34 (1995) 378-396.
- [14]. M. O. Gould, M. Kohlen, D. Keysers, H. Schubert, B. B. Wein, J. Bredno, T. M. Lehmann,
- [15]. Quality of DICOM header information for image categorization, in: International Symposium on Medical Imaging, Vol. 4685 of SPIE Proceedings, San Diego, CA, USA, 2002, pp. 280-287.
- [16]. C. LeBozec, M.-C. Jaulent, E. Zapletal, P. Degoulet, Unified modeling language and design of a case-based retrieval system in medical imaging, in: Proceedings of the Annual Symposium of the American Society for Medical Informatics (AMIA), Nashville, TN, USA, 1998.
- [17]. A. A. T. Bui, R. K. Taira, J. D. N. Dionision, D. R. Aberle, S. El-Saden, H. Kangarloo, Evidence-based radiology, Academic Radiology 9 (6) (2002) 662-669.

- [18]. J.-P. Boissel, M. Cucherat, E. Amsallem, P. Nony, M. Fardeheb, W. Manzi, M. C. Haugh, Getting evidence to prescribers and patients or how to make EBM a reality, in: Proceedings of the Medical Informatics Europe Conference (MIE 2003), St. Malo, France, 2003.
- [19]. C. E. Kahn, Artificial intelligence in radiology: Decision support systems, *Radiographic* 14 (1994) 849-861.
- [20]. H. Abe, H. MacMahon, R. Engelmann, Q. Li, J. Shiraishi, S. Katsuragawa, M. Aoyama, T. Ishida, K. Ashizawa, C. E. Metz, K. Doi, Computer-aided diagnosis in chest radiography: Results of large-scale observer tests at the 1996-2001 RSNA scientific assemblies, *Radio Graphics* 23 (1) (2003) 255-265.
- [21]. B. Kaplan, H. P. Lundsgaarde, Toward an evaluation of an integrated clinical imaging system: Identifying clinical benefits, *Methods of Information in Medicine* 35 (1996) 221-229.
- [22]. A. Horsch, R. Thurmayr, How to identify and assess tasks and challenges of medical image processing, in: Proceedings of the Medical Informatics Europe Conference (MIE 2003), St. Malo, France, 2003.
- [23]. S. Antani, L. R. Long, G. R. Thoma, A biomedical information system for combined content-based retrieval of spine x-ray images and associated text information, in: Proceedings of the 3rd Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP 2002), Ahmadabad, India, 2002.
- [24]. Müller H, Ruch, P, Geissbühler A., Enriching content-based medical image retrieval with automatically extracted MeSH.
- [25]. Juan C. Caicedo, Fabio A. Gonzalez and Eduardo Romero., Content-Based Medical Image Retrieval Using Low-Level Visual Features and Modality Identification.
- [26]. Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang., Image Retrieval: Ideas, Influences, and Trends of the New Age, The Pennsylvania State University.
- [27]. Giorgio Giacinto "A Nearest-Neighbor Approach to Relevance Feedback in Content Based Image Retrieval"
- [28]. Xiangli Xu, Libiao Zhang, Zhezhou Yu, Chunguang Zhou "Image Retrieval Using Multi-Granularity Color Features" ICALIP2008, IEEE.
- [29]. Joonwhan lee , eunjong park — Fuzzy similarity – based emotional classification of color images| vol 13,no.5,oct2011.
- [30]. yue gao,meng wang,zheng-jun zha — less is more efficient 3-D object retrieval with query view selection| vol 13, no.5,oct 2011
- [31]. yue gao,meng wang,zheng-jun zha — less is more efficient 3-D object retrieval with query view selection| vol 13, no.5,oct 2011
- [32]. T. M. Deserno, S. Antani, L. R. Long. Ontology of Gaps in Content-Based Image Retrieval. *J Digit Imaging*. February 2008. DOI:10.1007/s10278-007-9092-x.

Indian Two Wheeler Auto Industry and Concurrent Engineering

¹ Dr. V.Venkata Ramana ² V.Visweswara Reddy ³ T.Govardhan Reddy

¹ Professor, Mechanical Department, Ballari Institute of Technology and Management, INDIA

² Design Engineer, Hero Motocorp Ltd. (Sphinx Worldbiz Ltd.), Haryana, INDIA

³ Design Engineer, Infotech-Enterperises Ltd. Hyderabad, A.P, INDIA

ABSTRACT

In the wake of globalization, concurrent engineering (CE) is a promising connote for design, modification and development of new products for the challenging Indian auto industry. Success of CE demands that key areas of new product design and development of an organization need to be under constant focus. Most of the companies adopt CE procedures to reduce the time for introducing new product in to the market. This paper analyzes the impact of concurrent engineering practices followed by some selected Indian two wheeler automobile industries in the area of new product design and development as they seek to improve their competitive position in global markets. The research survey presents its evaluation based on analysis by application of statistical tools on the primary data which was collected through a well structured and pre-tested questionnaire. The results disclose that the selected two wheeler manufacturing companies in India are realizing maximum advantages with the implementation of the concurrent engineering in design and development of their new products.

KEYWORDS: Concurrent Engineering, New Product Development, Product and Process development, Statistical Tools.

I. INTRODUCTION

In the present global business scenario market is ever changing and volatile, resulting in shorter product life span. The firms must be able to take decisions on the spot instantly and responsively so as to reduce their product introduction time to market parallelly adapt to changing markets. Therefore, concurrent engineering (CE) has emerged as a way for rapid solutions in design and development process. The process of concurrent engineering is no doubt the mark of future in new product design and development for all companies regardless of their size, sophistication and to yield the gains, it requires a great deal of refinement in implementation of concurrent engineering process as it must be assessed and controlled continuously both in engineering and business processes. According to Clark and Fujimoto (1999)¹, "New product design and development is information and knowledge intensive work". Developing successful new products is possible through integration of abilities of both design and manufacturing expertise along with deriving firm's capabilities like ability to create, utilise and distribute knowledge throughout the process. Also knowledge sharing is unique and valuable resource for competitive advantage. Parsaei and Williams (2001)² state that CE is a product development methodology, which enhances productivity and lead to better overall designs relies strongly on the quality of information, interpretation, execution and implementation. CE is a "product or project approach where all activities of new product design and development operate simultaneously" and are closely coordinated to achieve optimal matching of requirements for effective cost, quality, and delivery. The automotive industry is the front-runner in many of the disciplines in their race to cut cost whilst remaining competitive. Smith (1998)³ discusses that the automotive firm's are seeking to take on major design responsibilities which need to significantly improve their ability to effectively conduct concurrent engineering early and often throughout the production process. Figure 1 shows a diagram representing the Concurrent Engineering (CE) Model.

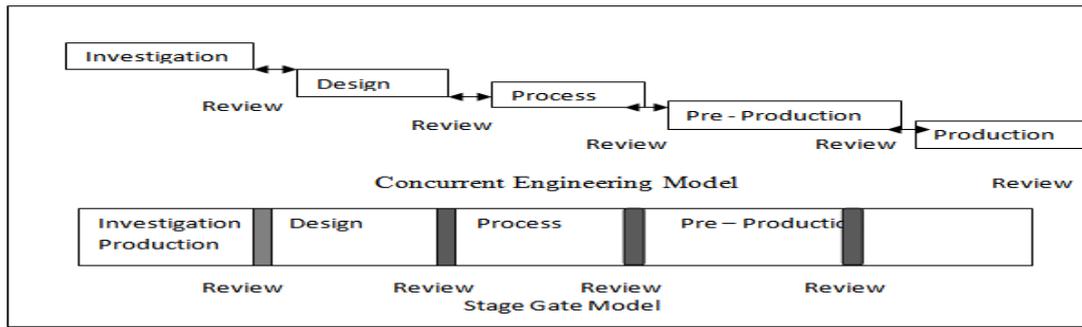


Figure 1: CE and Stage Gate Models

II. REVIEW OF LITERATURE

King N. and Majchrzak A. (1996)⁴ in his paper “The effect of project and process characteristics on product development cycle time” discussed about design integration as a “management process that integrates all activities from product concept through to production using multidisciplinary teams, to simultaneously optimize the product and its manufacturing processes to meet cost and performance objectives”. Griffin A. (2000)⁵ feels that design integration uses design tools such as modeling and simulation, teams and processes to develop products and their related processes concurrently. Design integration evolved in industry as an extension of work, such as Concurrent Engineering to improve customer satisfaction and competitiveness in a global economy. Verganti R. (1999)⁶ illustrates the importance of anticipating the capabilities of design integration during early development of the product. Early anticipation also referred to as forward planning means that information is anticipated as early as possible in the product development process so that solutions generated in the early phases already account for manufacturing constraints and opportunities. The challenge for global new product design and development with design integration is to achieve the prescribed activity whilst operating under the problems faced by teams are differences in communication and time, barriers between design, intermediate departments and customer lack of supplier involvement, management and teams working towards different goals and targets.

III. OBJECTIVE OF THE STUDY

The objective of the research study is to explore and establish the benefits gained by using concurrent engineering in product design and development on selected Indian two wheeler automobile industries.

IV. METHODOLOGY

The evaluation of the study is based on analysis of the primary data using Cumulative Weighted Average statistical analysis technique. The primary data was collected through a well-structure questionnaire from a sample size of 234 respondents of design, production and marketing groups of selected two wheeler manufacturing companies in India.

The questionnaire was sent to all respondents of 3 automobile companies and the usable response rate was 61% (see the Table 1)

Table 1: Response Rating of the Survey

Type of organization	Number of Organizations		Response Ratio (in percentage)
	Questionnaire Sent to	Response Received	
Two Wheeler Industry Automobile manufacturers	03	03	100%
No. of Respondents	234	143	61%

Data Analysis and Interpretation:

The data appropriate to the outcomes of concurrent engineering in new product design and development in two wheeler manufacturing companies are presented in the Table 2 and the same is depicted in the Figure 2.

Table 2: Impact of Concurrent Engineering

Variables	Cumulative Weighted Average
Use of concurrent engineering techniques	4.28
Encouraging external participation (suppliers and customers) in developing and designing the new products	3.76
Coordination of internal groups – design, Manufacturing	4.12
IT support and the knowledge management	3.82
Cross functional co-operation of multifunctional teams & Smoothing of organizational barriers	3.96
Ensuring Competitive Edge	4.06
Implementation of collaboration /partnership of managements	3.46
Management Reluctance	3.42
Average	3.8225

Source: Field Survey (Primary Data)

CWA: Cumulative Weighted Average

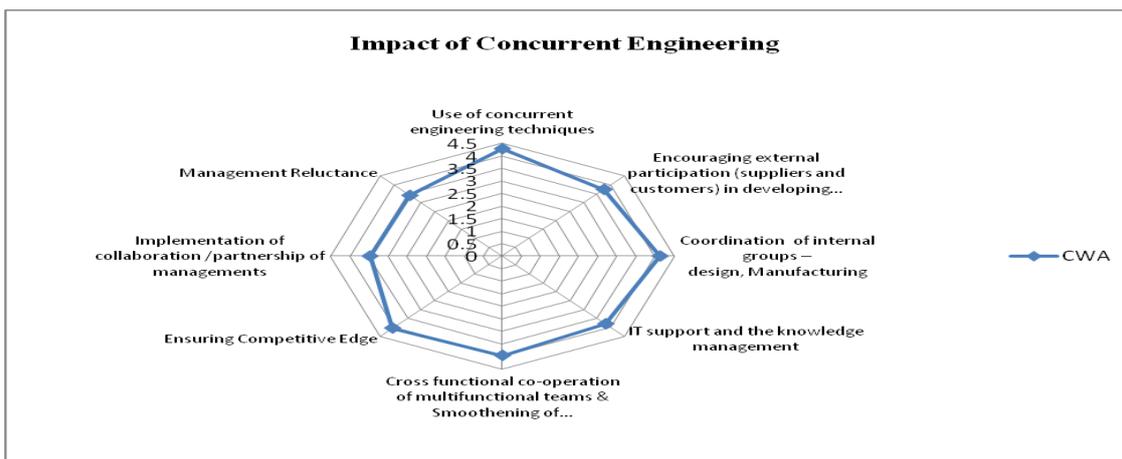


Figure 2: Impact of CE in Automobile industry

V. ELUCIDATION

The data presented in the Table 2, clearly establishes that the influence of concurrent engineering in new product design and development of two wheeler companies is considerably high in India. In specific, the variables with CWA score of greater than 3.75 on a 5-point scale: Use of concurrent engineering techniques, Encouraging external participation (suppliers and customers) in developing and designing the new products, Coordination of internal groups – design, manufacturing, IT support and the knowledge management, Cross functional co-operation of multifunctional teams & Smoothing of organizational barriers, Ensuring Competitive Edge, are identified as some of the key areas that realized most of the benefits.

VI. OUTCOMES AND DISCUSSION

Concurrent engineering has shown a very positive impact on design, development and introduction of new product in two wheeler automobile companies. Though companies are implementing concurrent engineering and realizing maximum benefits, they need to focus their attention in identifying the appropriate revolutionary technologies for proto-typing and thus increase cost savings and reduce time to market ultimately satisfying the customer needs.

VII. SCOPE FOR FUTURE WORK

The survey clearly reveals the facts that the CE has a great influence globally in new product design and development and stimulating the Indian auto sector currently. The survey made involved number of people related to the NPD, but ignored the expertise of the skilled people who are working in the concerned area but not qualified and the demographic factors such as gender and age were not taken into consideration which might influence the process to a greater extent. Further studies can be made by considering the above factors will reveal much more interesting facts. In addition the evaluation techniques such as ranking correlations can be applied to rank the different variables that are influencing the GNPD.

REFERENCES

- [1]. Clark K.B. and Fujimoto T. (1999) "*Product Development Performance Strategy Organization and Management in the World of the Auto Industry*", Boston Mass, Harvard Business School Press.
- [2]. Parsaei H.R and Williams G.S (2001) "*Concurrent Engineering, Contemporary issues and modern design tools*", Chapman and Hall, New York.
- [3]. Smith P.G. (1998) "*Developing Products in Half the Time*", 2E, John Wiley & Sons, Inc.
- [4]. King N. and Majchrzak A (1996) "*Concurrent Engineering Tools: Are the Human Issues Being Ignored?*" IEEE Transactions on Engineering Management, Vol. 43, No. 2, pp. 189-201.
- [5]. Griffin A, (2000), "*The effect of project and process characteristics on product development cycle time*". Journal of Marketing Research. Vol 34, issue 1, 24-35.
- [6]. Verganti R. (1999), "*Planned Flexibility: Linking Anticipation and Reaction in Product Development*" Projects. Journal of Product Innovation Management 16:363-376.
- [7]. Winner R.J., Pennel J.P., Bertrand H.E., Slusarczuk M.M., (1988), 'The role of concurrent engineering in weapons systems acquisition', IDA R-338, Institute for Defence Analyses.
- [8]. Shina, S.G (1991), "Concurrent Engineering: New Rules for World-Class Companies". Spectrum, IEEE Volume 28, Issue 7, Page(s):22 – 26.
- [9]. Rob Kinna, (1995), "Team Working and Concurrent Engineering – A Success Story", Journal: World Class Design to Manufacture, Volume: 2 Issue: 3 Page: 5 – 9.
- [10]. David Bradley, (1995), "Concurrent Engineering for Bespoke Products" Journal: Assembly Automation, Volume: 15 Issue: 1 Page: 35 – 37.
- [11]. Clark K.B. and Fujimoto T., (1991), 'Product Development Performance: Strategy, Organization and Management in the World Auto Industry', Boston: Harvard Business School Press, p.78.
- [12]. Ullrich K.T., Eppinger S.D., (1995), 'Product design and development', McGraw Hill, London.
- [13]. Schilling M.A., Hill C.W.L., (1998), 'Managing the new product development process: Strategic imperatives', Academy of Management Executive, vol. 12, no. 3.
- [14]. Clark K.B., Wheelwright S.C., (1993) "Managing new product development - *Text and Cases*", Harvard Business School.

Intrusion Detection In Manets Using Secure Leader Election

Yasmeen Shaikh, V. K. Parvati

Lecturer, Department of Computer Science and Engineering, VDRIT, Haliyal
HOD, Dept. of Information Science and Engineering, SDMCET, Dharwad

ABSTRACT

In this paper, we study leader election in the presence of selfish nodes for intrusion detection in mobile ad hoc networks (MANETs). To balance the resource consumption among all nodes and prolong the lifetime of a MANET, nodes with the most remaining resources should be elected as the leaders. However, there are two main obstacles in achieving this goal. First, without incentives for serving others, a node might behave selfishly by lying about its remaining resources and avoiding being elected. Second, electing an optimal collection of leaders to minimize the overall resource consumption may incur a prohibitive performance overhead, if such an election requires flooding the network. To address the issue of selfish nodes, we present a solution based on mechanism design theory. More specifically, the solution provides nodes with incentives in the form of reputations to encourage nodes in honestly participating in the election process. The amount of incentives is based on the Vickrey, Clarke, and Groves (VCG) model to ensure truth-telling to be the dominant strategy for any node. To address the optimal election issue, we propose an election algorithm that can lead to globally optimal election results with a low cost. We address these issues by assuming cluster of nodes. Finally, we justify the effectiveness of the proposed scheme through extensive experiments.

KEYWORDS : Leader election, intrusion detection systems, mechanism design and MANET security.

I. INTRODUCTION

The Mobile Ad hoc Networks (MANET) have no fixed chokepoints/bottlenecks where Intrusion Detection Systems (IDSs) can be deployed [3], [7]. Hence, a node may need to run its own IDS [14], [1] and cooperate with others to ensure security [15], [26]. This is very inefficient in terms of resource consumption since mobile nodes are energy-limited. To overcome this problem, a common approach is to divide the MANET into a set of one-hop clusters where each node belongs to at least one cluster. The nodes in each cluster elect a leader node (cluster head) to serve as the IDS for the entire cluster. The leader-IDS election process can be either random [16] or based on the connectivity [19]. Both approaches aim to reduce the overall resource consumption of IDSs in the network. However, we notice that nodes usually have different remaining resources at any given time, which should be taken into account by an election scheme. Unfortunately, with the random model, each node is equally likely to be elected regardless of its remaining resources. The connectivity index-based approach elects a node with a high degree of connectivity even though the node may have little resources left. With both election schemes, some nodes will die faster than others, leading to a loss in connectivity and potentially the partition of network. Although it is clearly desirable to balance the resource consumption of IDSs among nodes, this objective is difficult to achieve since the resource level is the private information of a node. Unless sufficient incentives are provided, nodes might misbehave by acting selfishly and lying about their resources level to not consume their resources for serving others while receiving others services. Moreover, even when all nodes can truthfully reveal their resource levels, it remains a challenging issue to elect an optimal collection of leaders to balance the overall resource consumption without flooding the network.

A Motivating Example

Figure 1 illustrates a MANET composed of ten nodes labeled from N1 to N10. These nodes are located in 5 one-hop clusters where nodes N5 and N9 belong to more than one cluster and have limited resources level. We assume that each node has different energy level, which is considered as private information. At this point, electing nodes N5 and N9 as leaders is clearly not desirable since losing them will cause a partition in the network and nodes will not be able to communicate with each other. However, with the random election model [16], nodes N5 and N9 will have equal probability, compared to others, in being elected as leaders. The nodes N5 and N9 will definitely be elected under the connectivity index-based approach due to their connectivity

indices [19]. Moreover, a naive approach for electing nodes with the most remaining resources will also fail since nodes' energy level is considered as private information and nodes might reveal fake information if that increases their own benefits. Finally, if the nodes N_2 , N_5 and N_9 are selfish and elected as leaders using the above models, they will refuse to run their IDS for serving others. The consequences of such a refusal will lead normal nodes to launch their IDS and thus die faster.

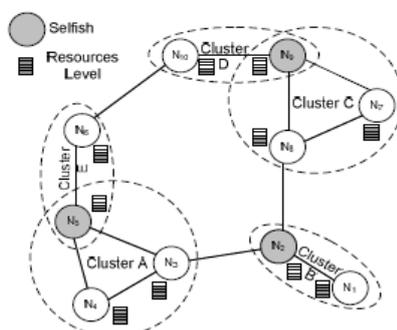


Figure 1: Example MANET

B Proposed system

In this paper, we propose a solution for balancing the resource consumption of IDSs among all nodes while preventing nodes from behaving selfishly. To address the selfish behavior, we design incentives in the form of reputation to encourage nodes to honestly participate in the election scheme by revealing their cost of analysis. The cost of analysis is designed to protect nodes' sensitive information (resources level) and ensure the contribution of every node on the election process (fairness). To motivate nodes in behaving normally in every election round, we relate the amount of detection service that each node is entitled to the nodes' reputation value. Besides, this reputation value can also be used to give routing priority and to build a trust environment.

The design of incentives is based on a classical mechanism design model, namely, Vickrey, Clarke, and Groves (VCG). The model guarantees that truth-telling is always the dominant strategy for every node during each election phase. On the other hand, to find the globally optimal cost-efficient leaders, a leader election algorithm is devised to handle the election process, taking into consideration the possibility of cheating and security flaws, such as replay attack. The algorithm decreases the percentage of leaders, single node clusters, maximum cluster size and increases average cluster size. The leaders are elected according to the received votes from the neighbor nodes, after the network is formulated into multiple clusters. Hence the leaders are elected in an optimal way in the sense that the resource consumption for serving as IDSs will be balanced among all nodes overtime. Finally, we justify the correctness of proposed methods through analysis and simulation. Empirical results indicate that our scheme can effectively improve the overall lifetime of a MANET.

The main contribution of this paper is a unified model that is able to:

1. Balance the IDS resource consumptions among all nodes by electing the most cost-efficient leaders.
2. Motivate selfish nodes to reveal their truthful resources level.

II. PROBLEM STATEMENT

We consider a MANET where each node has an IDS and a unique identity. To achieve the goal of electing the most cost efficient nodes as leaders in the presence of selfish and malicious nodes, the following challenges arise: First, the resource level that reflects the cost of analysis is considered as private information. As a result, the nodes can reveal fake information about their resources if that could increase their own benefits. Second, the nodes might behave normally during the election but then deviate from normal behavior by not offering the IDS service to their voted nodes.

In our model, we consider MANET as an undirected graph $G = (N, L)$ where N is the set of nodes and L is the set of bidirectional links. We denote the cost of analysis vector as $C = \{c_1, c_2, \dots, c_n\}$ where n is the number of nodes in N . We denote the election process as a function $v_{tk}(C, i)$ where $v_{tk}(C, i) = 1$ if a node i votes for a node k ; $v_{tk}(C, i) = 0$, otherwise. We assume that each elected leader allocates the same budget B (in the number of packets) for each node that has voted for it. Knowing that, the total budget will be distributed among all the voting nodes according to their reputation. This will motivate the nodes to cooperate in every election round that will be held on every time TELECT.

Thus, the model will be repeatable. The objective of minimizing the global cost of analysis while serving all the nodes can be expressed by the following Social Choice Function (SCF):

$$SCF = S(C) = \min \sum_{k \in N} c_k \cdot (\sum_{i \in N} vt_k(C, i) \cdot B)$$

Clearly, in order to minimize this SCF, the following must be achieved. First, we need to design incentives for encouraging each node in revealing its true cost of analysis value c , which will be addressed in Section III. Second, we need to design an election algorithm that can provably minimize the above SCF while not incurring too much of performance overhead. This will be addressed in Section V.

III. LEADER ELECTION MECHANISM

A The Mechanism Model

We treat the IDS resource consumption problem as a game where the N mobile nodes are the agents/players. Each node plays by revealing its own private information (cost of analysis) which is based on the node's type θ_i . The type θ_i is drawn from each player's available type set $\Theta_i = \{Normal, Selfish\}$. Each player selects his own strategy/type according to how much the node values the outcome. If the player's strategy is normal then the node reveals the true cost of analysis. We assume that each player i has a utility function [21]:

$$u_i(\theta_i) = p_i - v_i(\theta_i, \mathbf{o}(\theta_i, \theta_{-i})) \dots \dots \dots (1)$$

where,

- θ_{-i} is the type of all the other nodes except i .
- v_i is the valuation of player i of the output $\mathbf{o} \in O$, knowing that O is the set of possible outcomes. In our case, if the node is elected then v_i is the cost of analysis c_i . Otherwise v_i is 0 since the node will not be the leader and hence there will be no cost to run the IDS.
- $p_i \in \mathbb{R}$ is the payment given by the mechanism to the elected node. Payment is given in the form of reputation. Nodes that are not elected receive no payment.

B Cost of Analysis Function

During the design of the cost of analysis function, the following two problems arise: First, the energy level is considered as private and sensitive information and should not be disclosed publicly. Such a disclosure of information can be used maliciously for attacking the node with the least resources level. Second, if the cost of analysis function is designed only in terms of nodes' energy level, then the nodes with the low energy level will not be able to contribute and increase their reputation values.

To solve the above problems, we design the cost of analysis function with the following two properties: *Fairness* and *Privacy*. The former is to allow nodes with initially less resources to contribute and serve as leaders in order to increase their reputation. On the other hand, the latter is needed to avoid the malicious use of the resources level, which is considered as the most sensitive information. To avoid such attacks and to provide fairness, the cost of analysis is designed based on the reputation value, the expected number of time slots that a node wants to stay alive in a cluster and energy level. Note that the expected number of slots and energy level are considered as the nodes' private information.

To achieve our goal, we assume that the nodes are divided into l energy classes with different energy levels. The lifetime of a node can be divided into time-slots. Each node i is associated with an energy level, denoted by E_i , and the number of expected alive slots is denoted by nT_i . Based on these requirements, each node i has a power factor $PF_i = E_i/nT_i$. We introduce the set of $l - 1$ thresholds $P = \{\rho_1, \dots, \rho_{l-1}\}$ to categorize the classes as in Equation 2.

$$CL = \begin{cases} c_{l_1} & \text{if } PF < \rho_1 \\ c_{l_i} & \text{if } \rho_{i-1} \leq PF < \rho_i; i \in [2, l-1] \\ c_{l_l} & \text{if } PF \geq \rho_{l-1} \end{cases} \dots \dots \dots (2)$$

The reputation of node i is denoted by R_i . Every node has a sampling budget based on its reputation. The c_i notation represents the cost of analysis for a single packet and E_{ids} is used to express the energy needed to run the IDS for one time slot. The cost of analysis of each node can be calculated based on energy level. However, we considered energy level, expected lifetime and the present PS of node to calculate the cost of analysis. We can extend the cost of analysis function to more realistic settings by considering the computational level and cost of collecting and analyzing traffic. Our cost-of-analysis function is formulated as follows:

$$c_i = \begin{cases} \infty & \text{if } (E_i < E_{ids}) \\ \frac{PS_i}{PF_i} = \frac{R_i}{\frac{\sum_{t=1}^N R_t \times nT_t}{E_i}} & \text{otherwise} \end{cases}$$

Where PS is percentage of sampling defined as follows,

$$PS_i = \frac{R_i}{\sum_{t=1}^N R_t}$$

According to the above formulation, the nodes have an infinite cost of analysis if its remaining energy is less than the energy required to run the IDS for one time slot. This means that its remaining energy is too low to run the IDS for an entire time-slot. Otherwise, the cost of analysis is calculated through dividing the percentage of sampling by the power factor. The cost of analysis c is proportional to the percentage of sampling and is inversely proportional to the power factor. The rationale behind the definition of the function is the following. If the nodes have enough PS, they are not willing to loose their energy for running the IDS. On the other hand, if PF is larger, then the cost-of-analysis becomes smaller since the nodes have higher energy levels. In the rest of the paper, we will use cost and cost-of-analysis interchangeably. As the time goes by, the nodes belonging to lower energy class gains more budget while the budget of higher classes decreases. This justifies that our cost function is able to balance the energy of the nodes and gives a fair budget to all nodes.

C Reputation System Model

Before we design the payment, we need to show how the payment in the form of reputation can be used to: (1) Motivate nodes to behave normally and (2) punish the misbehaving nodes. Moreover, it can be used to determine whom to trust. To motivate the nodes in behaving normally in every election round, we relate the cluster's services to nodes' reputation. This will create a competition environment that motivates the nodes to behave normally by saying the truth. To enforce our mechanism, a punishment system is needed to prevent nodes from behaving selfishly after the election. Misbehaving nodes are punished by decreasing their reputation and consequently are excluded from the cluster services if the reputation is less than a predefined threshold. As an extension to our model, we can extend our reputation system to include different sources of information such as routing and key distribution with different assigned weights.

IV. IMPLEMENTATION

A Leader Election

To design the leader election algorithm, the following requirements are needed: (1) To protect all the nodes in a network, every node should be monitored by a leader. (2) To balance the resource consumption of IDS service, the overall cost of analysis for protecting the whole network is minimized. In other words, every node has to be affiliated with the most cost efficient leader among its neighbors. Our algorithm is executed in each node taking into consideration the following assumptions about the nodes and the network architecture:

- Every node knows its (2-hop) neighbors, which is reasonable since nodes usually maintain a table about their neighbors for routing purposes.
- Loosely synchronized clocks are available between nodes.
- Each node has a key (public, private) pair for establishing a secure communication between nodes.

To start a new election, the election algorithm uses four types of messages. *Hello*, used by every node to initiate the election process; *Begin-Election*, used to announce the cost of a node; *Vote*, sent by every node to elect a leader; *Acknowledge*, sent by the leader to broadcast its payment, and also as a confirmation of its leadership. For describing the algorithm, we use the following notation:

- **service-table(k)**: The list of all ordinary nodes, those voted for the leader node k .
- **reputation-table(k)**: The reputation table of node k . Each node keeps the record of reputation of all other nodes.
- **neighbors(k)**: The set of node k 's neighbors.
- **leadernode(k)**: The ID of node k 's leader. If node k is running its own IDS then the variable contains k .
- **leader(k)**: A boolean variable that sets to TRUE if node k is a leader and FALSE otherwise.

Initially, each node k starts the election procedure by broadcasting a *Hello* message to all the nodes that are one hop from node k and starts a timer $T1$. This message contains the hash value of the node's cost of analysis and its unique identifier (ID)

Algorithm 1 (Executed by every node)

/ On receiving Hello, all nodes reply with their cost */*

1. **if** (received *Hello* from all neighbors) **then**
2. Send *Begin-Election* ($ID_k, cost_k$);
3. **else if**(neighbors(k)= \emptyset) **then**
4. Launch IDS.
5. **end if**

On expiration of $T1$, each node k checks whether it has received all the hash values from its neighbors. Nodes from whom the *Hello* message has not received are excluded from the election. On receiving the *Hello* from all neighbors, each node sends *Begin-Election* as in Algorithm 1, which contains the cost of analysis of the node and then starts timer $T2$. If node k is the only node in the network or it does not have any neighbors then it launches its own IDS.

Algorithm 2 (Executed by every node)

/ Each node votes for one node among the neighbors */*

1. **if** ($\forall n _ neighbor(k), \exists i _ n : c_i \leq c_n$) **then**
2. send *Vote*($ID_k, ID_i, cost_{j=i}$);
3. *leadernode*(k):= i ;
4. **end if**

On expiration of $T2$, the node k compares the hash value of *Hello* to the value received by the *Begin-Election* to verify the cost of analysis for all the nodes. Then node k calculates the least-cost value among its neighbors and sends *Vote* for node i as in Algorithm 2. The *Vote* message contains the ID_k of the source node, the ID_i of the proposed leader and second least cost among the neighbors of the source node $cost_{j=i}$. Then node k sets node i as its leader in order to update later on its reputation. Note that the second least cost of analysis is needed by the leader node to calculate the payment. If node k has the least cost among all its neighbors then it votes for itself and starts timer $T3$.

Algorithm 3 (Executed by Elected leader node)

/ Send Acknowledge message to the neighbor nodes */*

1. *Leader*(i) := TRUE;
2. Compute Payment, P_i ;
3. *updateservice-table*(i);
4. *updatereputation-table*(i);
5. *Acknowledge* = P_i + all the votes;
6. Send *Acknowledge*(i);
7. Launch IDS.

On expiration of T3, the elected node *i* calculates its payment and sends an *Acknowledge* message to all the serving nodes as in Algorithm 3. The *Acknowledge* message contains the payment and all the votes the leader received. The leader then launches its IDS.

V. RESULTS AND DISCUSSIONS

A Simulation Environment

We simulate the scheme using Network Simulator 2(NS2). The main objective of our simulation result is to study the effect of node selection for IDS on the life of all nodes. To implement the model, we randomly assign 60 to 100 joules of energy to each node. We assume that the energy required for running the IDS for one time slot as 10 joules. We ignore the energy required to live and transmit packets to capture the silent aspect of the problem. We set the transmission radius of each node to 200 meters. Two nodes are considered as neighbors if their Euclidean distance is less than or equal to 200 meters. Besides, we deploy 20 nodes in an area of 500×500 square meters. It helps us to measure the performance of the nodes.

B Experimental Results

Nodes can behave selfishly before and after the election. A node shows selfishness before election by refusing to be a leader. On the other hand, selfishness after election is considered when nodes misbehave by not carrying out the detection service after being a leader. Both kinds of selfishness have serious impact on life of the normal nodes. As selfish nodes do not exhaust energy to run the IDS service, it will live longer than the normal nodes. In order to avoid this, election is based upon the reputation. Selfish nodes are punished with negative reputations. On other hand, node with highest reputation is elected as leader. Hence normal nodes will not launch IDS and will live longer. The nodes repetitively elect a set of leaders with maximum resources. Thus, we consider some nodes to be selfish and study their impact on our model. The experiment indicates that our model results in a higher percentage of alive nodes as shown in fig 2.

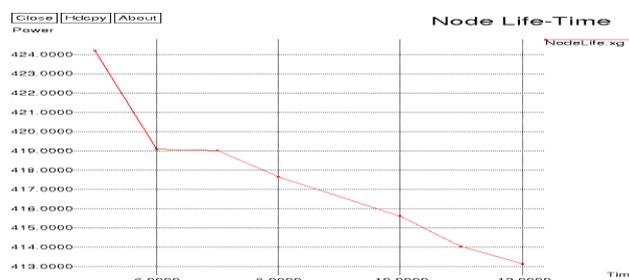


Figure 2. Node Life-Time

Our paper ensures that data can be transmitted in an easier, faster and secure way without intrusion. Leader election helps in identifying the intruders and hence data transmission rate is increased, thereby improving the throughput as shown in fig 3.

In our model, the node that has least cost of analysis becomes the leader. All nodes can keep a balance of their energy level with time. A leader has more energy hence capable of delivering large number of packets, thereby decreasing packet drop ratio and increasing packet delivery ratio as shown in fig 4 and 5 respectively.

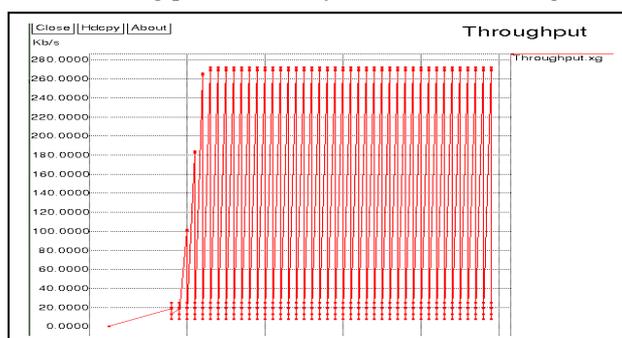


Fig 3 Throughput

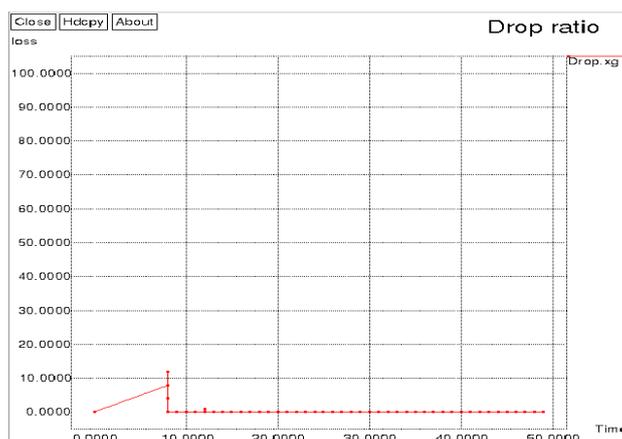


Figure 4 Packet drop ratio

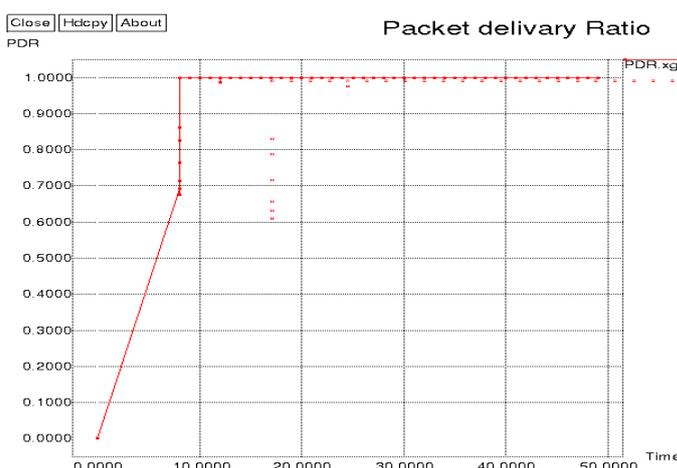


Figure 5 Packet Delivery ratio

VI. PERFORMANCE AND SECURITY ANALYSIS

A Performance

In this section, we analyze the performance overhead of our proposed leader algorithm. In summary, our algorithm has four steps. In the first 3 steps, all the nodes broadcast *Hello*, *Begin-Election* and *Vote* messages consecutively. In the last step, only the leader node sends an *Acknowledge* message to others.

1) *Computation Overhead*: Each node i signs its messages sent in the first 3 steps. Also, each node verifies the messages it received in these steps. In the 4th step, the leader node signs the *Acknowledge* message and others verify. Hence each normal node signs 3 messages and verifies $3|N_{gi}| + 1$ messages where N_{gi} is the number of neighboring nodes. On the other hand, the leader node signs 4 messages and verifies $3|N_{gi}|$ messages. Note that each node must find the least cost node which requires $O(\log(N_{gi}))$. Therefore, each node approximately performs $O(N_{gi})$ verifications, $O(1)$ signatures and $O(\log(N_{gi}))$ to calculate the least cost node. Thus the computation overhead for each node is $O(N_{gi}) + O(1) + O(\log(N_{gi})) \approx O(N_{gi})$. Since our algorithm involves more verification than signing, nodes can use the public key cryptosystem of [13] to verify a signature in 0.43s. Since leader election will take place after a certain interval, this computational overhead is tolerable.

2) *Communication Overhead*: Each node i broadcasts one message in the first 3 steps and only the leader node broadcasts a final *Acknowledge* message in the 4th step. Hence, the total communication overhead of our election algorithm is $3|N_{gi}| + 1 \approx O(N_{gi})$, where $|N_{gi}|$ is the number of neighboring nodes.

3) *Storage Overhead*: According to the algorithm, each node maintains a *reputation-table*, *neighbors* list and two variables: *Leadernode* and *leader*. The leader node keeps an extra *service-table*. Hence, each normal node needs $|N_i| + |N_{gi}| + 2$ storage and the leader node needs $|N_i| + |N_{gi}| + |V_i| + 2$. Knowing that $|N_i|$ is the number of nodes in the network, $|V_i|$ is the number of votes the leader node received where $|N_i| > |N_{gi}| > |V_i|$. Therefore, the total storage for each node is in the order of $O(N_i)$.

B Security

The main objective of our mechanism is to motivate selfish nodes and enforce them to behave normally during and after the election process. Here, we analyze the election mechanism in the presence of selfish and malicious nodes.

1. Presence of Selfish Nodes

A selfish node i will deviate from our mechanism if doing so increases its utility, u_i . Here we consider two type of untruthful revelation, namely, node i might either *under-declare* or *over-declare* the true value c_i of its cost of analysis. Node i may under-declare its valuation function with a fake value \hat{c}_i ($\hat{c}_i < c_i$). By under-declaring, node i pretends that it has a cheaper valuation function than reality. Since payments are designed based on VCG, playing by under-declaration will not help the node for two reasons. First, suppose the node i indeed has the lowest cost of analysis c_i , so it will win the election even by declaring its true value. In this case, reporting a lower value \hat{c}_i will not benefit the node because the payment is calculated based on the second best price and does not depend on the value declared by node i . Therefore, the utility of node i remains the same because it will be the difference between the payment and the real value c_i . Second, suppose that the node i does not have the cheapest valuation function but tries to win the election by revealing a lower value \hat{c}_i . This will help the node i to win the election but it will also lead to a negative utility function u_i for node i , because the payment it receives will be less than the real cost of analysis. That is, the node i will have to work more than what it has paid for. On the other hand, the node i might over-declare its valuation by revealing a fake \hat{c}_i ($\hat{c}_i > c_i$). Following such a strategy would never make a player happier in two cases. First, if the node i indeed has the cheapest valuation function, then following this strategy may prevent the node from being elected, and therefore it will lose the payment. On the other hand, if node i still wins, then its utility remains the same since the payment does not depend on the value it reports. Second, suppose the real valuation function c_i of node i is not the lowest, then reporting a higher value will never help the node to win. Last but not least, the checkers are able to catch and punish the misbehaving leaders by mirroring a portion of its computation from time to time. A caught misbehaving leader will be punished by receiving a negative payment. Thus it discourages any elected node from not carrying out its responsibility. We can thus conclude that our mechanism is truthful and it guarantees a fair election of the most cost efficient leader.

2. Presence of Malicious Nodes

A malicious node can disrupt our election algorithm by claiming a fake low cost in order to be elected as a leader. Once elected, the node does not provide IDS services, which eases the job of intruders. Due to the presence of checkers, a malicious node has no incentive to become a leader. Since it will be caught and punished by the checkers. After a leader is caught misbehaving, it will be punished by receiving a negative reputation and is consequently excluded from future services of the cluster. Thus, our mechanism is still valid even in the presence of a malicious node.

VII. ADVANTAGES AND DISADVANTAGES

A. Advantages

- Ensures easier, secure and faster data transmission.
- Balanced resource consumption in the presence of selfish and malicious nodes.
- Decreases the percentage of leaders, single node clusters, maximum cluster size and increase average cluster size.
- Life time of nodes is increased as nodes with maximum resources are elected as leaders.
- Leaders are elected randomly based upon the reputation.

B. DISADVANTAGES

- If any new node is added or removed from the network, it is not possible to dynamically configure the network, which may lead to partitioning in the network.
- It is not possible to make a specified node as leader who would launch IDS, as leaders are elected randomly.

VIII. FUTURE SCOPE

In future work, we plan expanding our model to dynamically configure the network when any new node is created or when any existing nodes is removed from the network. We'll also plan to make a specified node as leader by manipulating the leader election mechanism.

IX. CONCLUSION

The unbalanced resource consumption of IDSs in MANET and the presence of selfish nodes have motivated us to propose an integrated solution for prolonging the lifetime of mobile nodes and for preventing the emergence of selfish nodes. The solution motivated nodes to truthfully elect the most cost efficient nodes that handle the detection duty on behalf of others. Moreover, the sum of the elected leaders is globally optimal. To achieve this goal, incentives are given in the form of reputations to motivate nodes in revealing truthfully their costs of analysis. Reputations are computed using the well known VCG mechanism by which truth-telling is the dominant strategy. We also analyzed the performance of the mechanisms in the presence of selfish and malicious nodes. To implement our mechanism, we devised an election algorithm with reasonable performance overheads. Simulation results showed that our model is able to prolong the lifetime and balance the overall resource consumptions among all the nodes in the network. Moreover, we are able to decrease the percentage of leaders, single node clusters, maximum cluster size and increase average cluster size. These properties allow us to improve the detection service through distributing the sampling budget over less number of nodes and reduce single nodes to launch their IDS.

REFERENCES

- [1] T. Anantvalee and J. Wu. A survey on intrusion detection in mobile adhoc networks. *Wireless/Mobile Network Security*, 2006.
- [2] L. Anderegg and S. Eidenbenz. Ad hoc-VCG: A truthful and cost efficient routing protocol for mobile ad hoc networks with selfish agents. In *proc. of the ACM International Conference on Mobile Computing and Networking (MobiCom)*, 2003.
- [3] F. Anjum and P. Mouchtaris. *Security for Wireless Ad Hoc Networks*. John Wiley & Sons. Inc., USA, 2007.
- [4] S. Basagni. Distributed and mobility-adaptive clustering for multimedia support in multi-hop wireless networks. In *proc. of the IEEE International Vehicular Technology Conference (VTC)*, 1999.
- [5] S. Basagni. Distributed clustering for ad hoc networks. In *proc. of the IEEE International Symposium on Parallel Architectures, Algorithms, and Networks (ISPAN)*, 1999.
- [6] M. Bechler, H. Hof, D. Kraft, F. Pahlke, and L. Wolf. A cluster based security architecture for ad hoc networks. In *proc. of the IEEE INFOCOM*, 2004.
- [7] P. Brutch and C. Ko. Challenges in intrusion detection for wireless adhoc networks. In *proc. of the IEEE Symposium on Applications and the Internet (SAINT) Workshop*, 2003.
- [8] S. Buchegger and J. L. Boudec. Performance analysis of the CONFIDANT protocol (cooperation of nodes - fairness in dynamic adhoc networks). In *proc. of the ACM MOBIHOC*, 2002.
- [9] K. Chen and K. Nahrstedt. iPass: An incentive compatible auction scheme to enable packet forwarding service in MANET. In *proc. of the International Conference on Distributed Computing Systems*, 2004.
- [10] B. DeCleene, L. Dondeti, S. Griffin, T. Hardjono, D. Kiwior, J. Kurose, D. Towsley, S. Vasudevan, and C. Zhang. Secure group communications for wireless networks. In *proc. of the IEEE Military Communications Conference (MILCOM)*, 2001.
- [11] J. Feigenbaum, C. Papadimitriou, R. Sami, and S. Shenker. A BGP based mechanism for lowest-cost routing. In *proc. of the ACM symposium on Principles of distributed computing (PODC)*, 2002.
- [12] J. Feigenbaum and S. Shenker. Distributed algorithmic mechanism design: Recent results and future directions. In *proc. of the AMM International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications (DIALM)*, 2002.
- [13] N. Gura, A. Patel, A. Wander, H. Eberle, and S. C. Shantz. Comparing elliptic curve cryptography and RSA on 8-bit CPUs. In *proc. of the Cryptographic Hardware and Embedded Systems (CHES)*, 2004.
- [14] S. Gwalani, K. Srinivasan, G. Vigna, E. M. Beding-Royer, and R. Kemmerer. An intrusion detection tool for AODV-based ad hoc wireless networks. In *proc. of the IEEE Computer Security Applications Conference (CSAC)*, 2004.
- [15] Y. Hu, A. Perrig, and D. B. Johnson. Ariadne: A secure on-demand routing protocol for ad hoc networks. In *proc. of the ACM International Conference on Mobile Computing and Networking (MOBICOM)*, 2002.
- [16] Y. Huang and W. Lee. A cooperative intrusion detection system for adhoc networks. In *proc. of the ACM Workshop on Security of Ad Hoc and Sensor Networks*, 2003.
- [17] L. Hurwicz and S. Reiter. *Designing Economic Mechanisms*. Cambridge University Press, 1st edition, 2008.
- [18] J. Green and J. Laffont. *Incentives in Public Decision-Making*. Springer Netherlands, USA, 1996.
- [19] O. Kachirski and R. Guha. Efficient intrusion detection using multiple sensors in wireless ad hoc networks. In *proc. of the IEEE Hawaii International Conference on System Sciences (HICSS)*, 2003.
- [20] P. Krishna, N. H. Vaidya, M. Chatterjee, and D. K. Pradhan. A cluster based approach for routing in dynamic networks. In *proc. of the ACM SIGCOMM Computer Communication Review*, 1997.
- [21] A. Mas-Colell, M. Whinston, and J. Green. *Microeconomic Theory*. Oxford University Press, New York, 1995.
- [22] P. Michiardi and R. Molva. Analysis of coalition formation and cooperation strategies in mobile adhoc networks. *Journal of Ad hoc Networks*, 3(2):193 – 219, 2005.
- [23] A. Mishra, K. Nadkarni, and A. Patcha. Intrusion detection in wireless ad hoc networks. *IEEE Wireless Communications*, 11(1):48 – 60, 2004.
- [24] N. Mohammed, H. Otrok, L. Wang, M. Debbabi, and P. Bhattacharya. A mechanism design-based multi-leader election scheme for intrusion detection in manet. In *proc. of the IEEE Wireless Communications & Networking Conference (WCNC)*, 2008.
- [25] P. Morris. *Introduction to Game Theory*. Springer, 1st edition, 1994.
- [26] P. Ning and K. Sun. How to misuse AODV: A case study of insider attacks against mobile ad-hoc routing protocols. In *proc. of the IEEE Information Assurance Workshop*, 2003.
- [27] N. Nisan and A. Ronen. Algorithmic mechanism design. In *Games and Economic Behavior*, pages 129–140, 1999.
- [28] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, 1st edition, 2007.

- [29] H. Otrok, N. Mohammed, L. Wang, M. Debbabi, and P. Bhattacharya. A game-theoretic intrusion detection model for mobile ad-hoc networks. *Journal of Computer Communications*, 31(4):708 – 721, 2008.
- [30] A. Perrig, R. Canetti, D. Tygar, and D. Song. The TESLA broadcast authentication protocol. *RSA Cryptobytes*, 5(2):2 – 13, 2002.
- [31] J. Shneidman and D. Parkes. Specification faithfulness in networks with rational nodes. In *proc. of the ACM Symposium on Principles of Distributed Computing*, 2004.
- [32] K. Sun, P. Peng, P. Ning, and C. Wang. Secure distributed cluster formation in wireless sensor networks. In *proc. of the IEEE Computer Security Applications Conference (ACSAC)*, 2006.
- [33] S. Vasudevan, B. DeCleene, N. Immerman, J. Kurose, and D. Towsley. Leader election algorithms for wireless ad hoc networks. In *proc. of the IEEE DARPA Information Survivability Conference and Exposition (DISCEX III)*, 2003.
- [34] S. Vasudevan, J. Kurose, and D. Towsley. Design and analysis of a leader election algorithm for mobile ad hoc networks. In *proc. of the IEEE International Conference on Network Protocols (ICNP)*, 2004.
- [35] Y. Zhang and W. Lee. Intrusion detection in wireless ad-hoc networks. In *proc. of the ACM International Conference on Mobile Computing and Networking (MobiCom)*, 2000.

Involute Gear Tooth Contact And Bending Stress Analysis

Vishwjeet V.Ambade¹, Prof. Dr. A.V.Vanalkar², Prof.P.R.Gajbhiye³

¹ P.G Student , K.D.K College Of Engineering, Nagpur, India

^{2,3}. Faculty, Kdk College Of Engineering Nagpur, India

ABSTRACT

This paper presented analysis of Bending stress and Contact stress of Involute spur gear teeth in meshing. There are several kinds of stresses present in loaded and rotating gear teeth. Bending stress and contact stress (Hertz stress) calculation is the basic of stress analysis. It is difficult to get correct answer on gear tooth stress by implying fundamental stress equation, such as Lewis formula for bending stress and Hertz equation for contact stress. Various research methods such as Theoretical, Numerical and Experimental have been done throughout the years. This paper shows the theoretical and numerical approach to calculate bending and contact stress. The results were further compared with ANSYS result to validate.

KEYWORDS: Spur Gear, Bending Stress, Contact Stress, ANSYS.

I. INTRODUCTION

When one investigates actual gears in service, the conditions of the surface and bending failure are two of the most important features to be considered. The finite element method is very often used to analyze the stress states of elastic bodies with complicated geometries, such as gears. There are published papers, which have calculated the elastic stress distributions in gears. In these works, various calculation methods for the analysis of elastic contact problems have been presented. The finite element method for two-dimensional analysis is used very often. It is essential to use a three-dimensional analysis if gear pairs are under partial and non uniform contact. However, in the three dimensional calculation, a problem is created due to the large computer memory space that is necessary. In this chapter to get the gear contact stress a 2-D model was used. Because it is a nonlinear problem it is better to keep the number of nodes and elements as low as possible. In the bending stress analysis -D models are used for simulation.

II. ANALYTICAL PROCEDURE

From the results obtained in chapter 3 the present method is an effective and accurate method, which is proposed to estimate the tooth contact stresses of a gear pair.. Using the present method, the tooth contact stresses and the tooth deflections of a pair of spur gears analyzed by ANSYS 7.1 Since the present method is a general one, it is applicable to many types of gears. In early works, the following conditions were assumed in advance:

- There is no sliding in the contact zone between the two bodies
- The contact surface is continuous and smooth

Using the present method ANSYS can solve the contact problem and not be limited by the above two conditions. A two-dimensional and an asymmetric contact model were built. First, parameter definitions were given and then many points of the involute profile of the pinion and gear were calculated to plot an involute profile using a cylindrical system. The equations of an involute curve below were taken from Buckingham.

$$r = r_b * (1 + \beta^2)^{1/2}$$

$$\theta = \tan\phi - \phi = \text{inv}\phi$$

where r = radius to the involute form, $b r$ = radius of the base circle

$$\beta = \phi + \xi$$

θ = vectorial angle at the pitch circle

ξ = vectorial angle at the top of the tooth

ϕ = pressure angle at the pitch circle

ϕ = pressure angle at radius r

One spur tooth profile was created using equation , shown in Figure 1, as are the outside diameter circle, the dedendum circle, and base circle of the gear. Secondly, in ANSYS from the tool bars using “CREATE”, “COPY”, “MOVE”, and “MESH” and so on, any number of teeth can be created and then kept as the pair of gear teeth in contact along the line of the action. The contact conditions of gear teeth are sensitive to the geometry of the contacting surfaces, which means that the element near the contact zone needs to be refined. It is not recommended to have a fine mesh everywhere in the model, in order to reduce the computational requirements. There are two ways to build the fine mesh near the contact surfaces. One is the same method as presented in chapter 3, a fine mesh of rectangular shapes were constructed only in the contact areas. The other one, “SMART SIZE” in ANSYS, was chosen and the fine mesh near the contact area was automatically created.

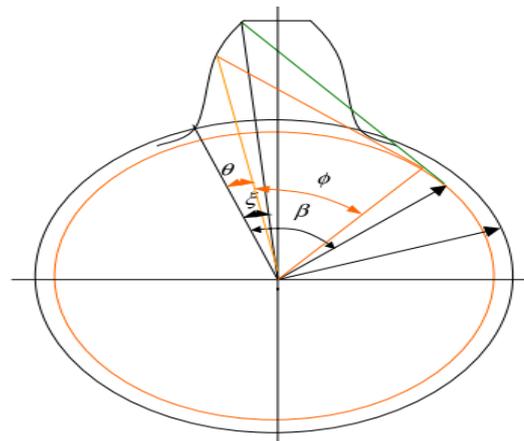


Figure 1 - Involutometry of a spur gear

III. ROTATION COMPATIBILITY OF THE GEAR BODY

In order to know how much load is applied on the contact stress model and the bending stress model, evaluating load sharing between meshing gears is necessary. It is also an important concept for transmission error. It is a complex process when more than one-tooth pair is simultaneously in contact taking into account the composite tooth deflections due to bending, shearing and contact deformation. This section presents a general approach as to how the load is shared between the meshing teeth in spur gear pairs. When the gears are put into mesh, the line tangent to both base circles is defined as the line of action for involute gears. In one complete tooth mesh cycle, the contact starts at points A shown in Figure 2 where the outside diameter circle, the addendum circle of the gear intersects the line of action. The mesh cycle ends at point E, as shown in Figure 3 where the outside diameter of the pinion intersects the line of action.

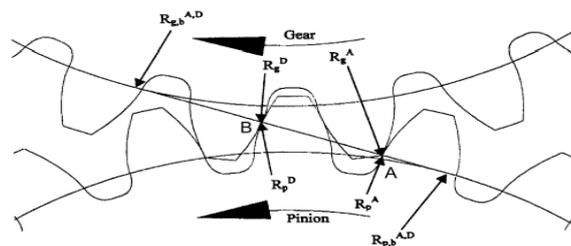


Figure 2 Illustration of one complete tooth meshing cycle

Consider two identical spur gears in mesh. When the first tooth pair is in contact at point A it is between the tooth tip of the output gear and the tooth root of the input gear (pinion). At the same time a second tooth pair is already in contact at point D in Figure 4.3. As the gear rotates, the point of contact will move along the line of action APE. When the first tooth pair reaches point B shown in Figure 4.4, the second tooth pair disengage at point E leaving only the first tooth pair in the single contact zone. After this time there is one pair of gear in contact until the third tooth pair achieves in contact at point A again. When this tooth pair rotates to point D, the another tooth pair begins engagement at point A which starts another mesh cycle. After this time there are two pairs of gear in contact until the first tooth pair disengage at point

E. Finally, one complete tooth meshing cycle is completed when this tooth pair rotates to point E. To simplify the complexity of the problem, the load sharing compatibility condition is based on the assumption that the sum of the torque contributions of each meshing tooth pair must equal the total applied torque.

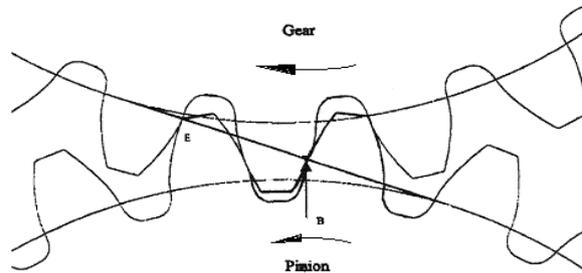


Figure 3 Different positions for one complete tooth meshing cycle

Analytical equations can also be developed for the rotation of the gear and pinion hubs, including the effects of tooth bending deflection and shearing displacement and contact deformation. In the pinion reference frame, it is assumed that the pinion hub remains stationary, while the gear rotates due to an applied torque. Considering the single pair contact zone at point B, the condition of angular rotation of the gear body will then be given by

For the pinion

For the gear

where and are the tooth displacement vectors caused by bending and shearing for pairs B of the pinion and gear respectively,

and are the contact deformation vectors of tooth pair B of the pinion and gear respectively.

denotes the transverse plane angular rotation of the pinion body caused by bending deflection, shearing displacement and contact deformation of the tooth pair B while the gear is stationary. Conversely, for the gear rotation while the pinion is stationary, gives the transverse plane angular rotations of the gear body.

IV. CALCULATION PART OF CONTACT STRESS, BENDING STRESS AND DEFORMATION.

Different Terms used in Calculation of Deformation and Bending Stress.

Velocity Factor between Gear and Pinion (K_v) = 1

Face width (F) = 40mm.

Module (m) = 01 (Between gear and pinion).

No. of teeth (x), Gear=40 and Pinion=20

Pitch radius Pinion (r_{p1}) = 50mm

Pitch radius Gear (r_{p2}) = 95mm

Pressure Angle (ϕ) = 20°

Constant (K_1) = 01

Constant (K_2) = 01

Total Load (W) = 1000N.

Poisson's Ratio (μ) = 0.3

y = 20mm.

Y = 10mm.

Addendum of Gear (h_1) and Pinion (h_2) = 10mm.

Young Modulus of Elasticity = 2×10^5

Tooth Thickness = 14.95mm.

Gear Ratio = 2.

Bending Stress Calculation.

$$\sigma_b = \frac{W_t}{b \times m \times y}$$

Where,

$$W_t =$$

mm

$$\sigma_b = \frac{W_t}{K_v \times F \times m \times y}$$

$$\sigma_b = \frac{1000}{1 \times 40 \times 1 \times 26.66} = 0.93 \text{ N/mm}^2$$

Contact Stress (Hertz Stress)

Width of the band of contact (B).

$$B = \sqrt{\frac{16 \times W (K_1 + K_2) R_1 \times R_2}{l (R_1 + R_2)}}$$

$$B = \sqrt{\frac{16 \times 1000 (1 + 1) 17.10 \times 32.49}{84.44(17.10 + 32.49)}} = 46.07 \text{ mm}$$

Where R₁, R₂= Radius of Curvature of an involutes curves at the contact points.

$$R_1 = r_{p1} \sin \phi = 50 \sin 20 = 17.10$$

$$R_2 = r_{p2} \sin \phi = 95 \sin 20 = 32.49$$

$$l^2 = \frac{2 k W R}{\pi} = 7130.14 \text{ mm} \quad \text{and} \quad l = 84.44 \text{ mm}$$

$$R = \left(\frac{1}{R_1} - \frac{1}{R_2} \right)^{-1} = \left(\frac{1}{17.10} - \frac{1}{32.49} \right)^{-1} = 11.20 \text{ mm}$$

Maximum Contact Stress (

$$\sigma_c = \frac{4 \times W}{L \times \pi \times B}$$

$$\sigma_c = \frac{4 \times 1000}{84.44 \times \pi \times 46.07} = 0.32 \text{ N/mm}^2$$

To Determine Deformation (δ)

Coefficient of Compliance (C) .

$$C = \frac{E \times b \times \delta}{P}$$

$$C = C_B + C_F + C_H$$

Where:

C_B= Bending Compliance.

C_H= Hertz Compliance.

C_F= Root Compliance.

$$C_B = 12 \times \cos^2 \phi \left[I_1 + I_2 \left\{ 0.2(1 + \mu) + \frac{\tan^2 \phi}{12} \right\} \right]$$

$$I_1 = \int_0^y \frac{dy}{t} = \int_0^{20} \frac{20}{17} dy = 23.52$$

$$I_2 = \int_0^y \frac{(y - Y)^2}{t^3} \times dy = \int_0^{20} \frac{(10 - 20)^2}{17^3} dy = 0.40$$

$$C_B = 12 \times \cos^2(20) \left[23.52 + 0.40 \left\{ 0.2(1 + 0.3) + \frac{\tan^2 20}{12} \right\} \right] = 65.78$$

$$C_H = \frac{2(1 - \mu^2)}{\pi} \left[In \frac{4h_1 h_2}{l^2} - \frac{\mu}{1 - \mu} \right]$$

$$C_H = \frac{2(1 - 0.3^2)}{\pi} \left[In \frac{4 \times 10 \times 10}{7130.4} - \frac{0.3}{1 - 0.3} \right] = -0.95$$

Where

are the addendum of gear and pinion = 10mm.

$$C_F = (1 - \mu)^2 \cos^2 \phi \left[\frac{50}{3\pi} \left(\frac{y}{t_f} \right)^2 + \frac{2(1 - 2\mu)}{1 - \mu} \times \frac{y}{t_f} + \frac{4.82}{\pi} \left(1 + \frac{\tan^2 \phi}{2.4(1 + \mu)} \right) \right]$$

$$C_F = (1 - 0.3)^2 \cos^2 20 \left[\frac{50}{3\pi} \left(\frac{20}{18} \right)^2 + \frac{2(1 - 2 \times 0.3)}{1 - 0.3} \times \frac{20}{18} + \frac{4.82}{\pi} \left(1 + \frac{\tan^2 20}{2.4(1 + .3)} \right) \right]$$

$$C_F = 3.096$$

$$C = C_B + C_F + C_H$$

$$C = 65.78 + 3.096 + (-0.95) = 67.926$$

$$C = \frac{E \times b \times \delta}{P}$$

$$67.926 = \frac{2 \times 10^9 \times 40 \times \delta}{1000}$$

Contact Stress, bending stress and deformation where further validate by FEM Analysis. Reference

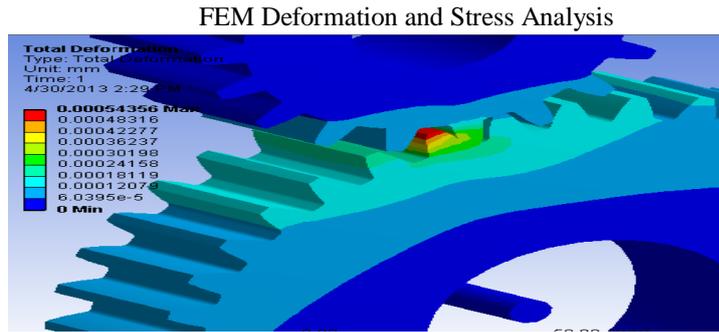


Fig-4 Deformation at 1 KN

Deformation Analysis: It has been found that deformation is maximum at the top land of face width. This indicates that top land area of spur gear is under high temperature and pressure due to which deformation is maximum at its top land. Here the deformation is 0.00054mm and calculated deformation is 0.00084mm. by comparing this two results, we can say that calculated deformation and Analytical deformation is validate.

Bending Stress Analysis:

It has been found that Stress is maximum at the contact surface of mating gear as shown in fig-5. The calculated bending is 0.93 MPa and contact stress is 0.32 MPa. So the calculated equivalent stress can be consider 1.25 MPa. But the Analytical Equivalent stress is found to be 4.416 MPa. So here the further modification is required to validate the results.

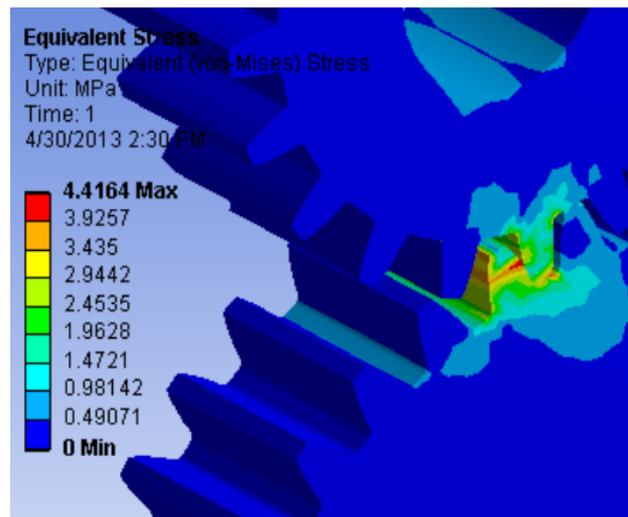


Fig 5 Stress at 1KN

REFERENCE

- [1] K Ruthupavam & Y Sandeep Kumar (2012) "Optimization of Design Based on Tip Radius and Tooth Width to Minimize the Stresses on the Spur Gear with FE Analysis". HTC 2012
- [2] Manoj Hariharan (2006) "spur gear tooth stress analysis and stress reduction using stress reducing geometrical features" Thesis Report Submitted to Thapar Institute of Engineering and Technology
- [3] Namam M. Ahmed "Stress Distribution along the Involute Curve of Spur Gears" Thesis Report Submitted to Institute of Technology, Sulamani Iraq.
- [4] W.H.Dornfeld (2004) "Gear Tooth Strength Analysis" Book.
- [5] Zeping wei (2004) "stresses and deformations in involute spur gears by finite element method" Thesis Report Submitted to University of Saskatchewan.
- [6] F. K.Gopinath & M.M.mayuram "spur gear - tooth stresses" Book
- [7] Rixin Xu (2008) "finite element modeling and simulation on the quenching effect for spur gear design optimization" Thesis Report Submitted to University of Akron
- [8] kristina marković – marina franulović (2011) "contact stresses in gear teeth due to tip relief profile modification" Eng. Rev. 31-1 (2011) 19-26
- [9] Chuen-Huei Liou and Hsiang Hsi Lin (1992) "Effect of Contact Ratio on Spur Gear Dynamic Load" NASA AVSCOM Technical Memorandum 105606 Technical Report 91-C-025
- [10] Ashwini Joshi, Vijay Kumar Karma (2011) "Effect on Strength of Involute Spur Gear by Changing the Fillet Radius Using FEA" International Journal Of Scientific & Engineering Research Volume2, Issue 9, September-2011, ISSN 2229-5518

- [11] GD. Bibel, S.K. Reddy, and M. Savage (1991) "Effects of Rim Thickness on Spur Gear Bending Stress" NASA III0wUIIII AVSCOM Technical Memorandum 104388 Technical Report 91-C-015
- [12] Sweta Nayak and Swetleena Mishra (2007) "effects of addendum modification on root stress in involute spur gears" Thesis Report Submitted to National Institute of Technology Rourkela
- [13] Sorin Cananau(2003) "3d contact stress analysis for spur gears" national tribology conference 24-26 september 2003
- [14] G.Mallesh (2009) "Effect of Tooth Profile Modification In Asymmetric Spur Gear Tooth Bending Stress By Finite Element Analysis" NaCoMM-2009- ASMG18268
- [15] Dr. Ir H.G. H. van Melick (2007) "Tooth-Bending Effects in Plastic Spur Gears Influence on load sharing, stresses and wear, studied by FEA" Gear Technology Pune-2007.
- [16] M Koilraj, Dr G Muthuveerappan and Dr J Pattabi-raman, "An Improvement in Gear Tooth Design Methodology using Finite Element Method", IE(I) Journal MC, Volume 88, October 2007.

Model based numerical state feedback control of jet impingement cooling of a steel plate by pole placement technique

Purna Chandra Mishra¹, Swarup Kumar Nayak¹, Santosh K. Nayak¹,

Premananda Pradhan¹, Durga P. Ghosh¹

¹School of Mechanical Engineering, KIIT University, Patia, Bhubaneswar- 751024, Odisha, India

ABSTRACT

This article presents the application of a state feedback, Pole-Placement temperature-tracking control formulation leading to a comprehensive Single-input, multiple-output (SIMO) structure for cooling of a steel plate by an impinging air jet. Following a semi-discrete control volume formulation of the 1-D transient heat conduction; the state space model of nodal variation of metal temperature with time has been arrived. Stable temperature track with respect to the reference temperature track was obtained by using the state feedback control algorithm. To solve this prescribed state feedback problem, an efficient pole placement technique was implemented in two ways i.e., initially, the fixed poles and then variable poles in the complex plane. The control models were simulated in the MATLAB and SIMULINK environments. The state feedback pole placement technique was found efficient in controlling the parameters in the given application. It was observed that the response of the non-linear system is sensitive to linearization time interval. Better control is implemented by increasing the frequency of adjustment of the closed-loop poles.

KEYWORDS: Feedback Control, Jet Impingement Cooling, Pole Placement Technique, State Space Method, Single Input Multiple Output Problem, Fixed Poles, Variable Poles, Temperature Control

Nomenclature:

K: The feedback gain matrix
 A: The coefficient matrix in the state space equation
 A_i: The coefficient matrix in the state space representation for ith model
 \bar{A} : The mean coefficient matrix
 B: The coefficient matrix of control input
 T₁: Temperature at the top surface of the steel plate (K)
 B_i: The coefficient matrix of control input or ith model
 X: The matrix of state variables
 \bar{X} : The matrix of mean state variables
 \dot{X} : The matrix of state rates
 a_c: The perturbation in strain rates
 X' : The error matrix $(X - \bar{X})$
 λ : The eigen values
 \dot{T}_1 : The demand (K/s)
 P_i (S): The ith number of transformable models
 a: The control input (Strain rate)
 \bar{a} : The mean strain rate (mean control output)
 A-BK: The feedback matrix
 T: The transformation matrix
 I: The identity matrix
 Z: The assumed state variable matrix
 r: The number of operating conditions
 t: Time (s)

I. INTRODUCTION

Heat treatment involves the method of controlled heating, soaking and cooling temperature cycles that are applied to a material or substrate, to impart superior metallurgical properties to it. Heat treatment has wide application and is most abundantly used in the field of glass, ceramic, iron and steel production industries.

The dynamics of heat transfer systems are highly nonlinear. Irrespective of nonlinearity in of heat transfer systems, air jet impingement cooling systems also have model uncertainties which includes large changes in the air flow rate and variation of parameters, the external disturbances, leakages, and friction. To overcome this problem in order to gain profits, process designers are creating designs and modelling in complex non-linearity regions where process controllers can face stiff the challenges. In model predictive control, the controller action is the solution to a constrained optimization problem that is solved numerically on-line [1, 2, 3]. In contrast, differential geometric control is a direct synthesis approach in which the controller is derived by requesting a desired closed loop response in the absence of input constraints [4, 5]. In Lyapunov-based control, closed loop stability plays a central basic role in a controller design. [6, 7, 8]. Kravaris and Daoutidis [9] presented a non-linear state feedback controller for second order non-linear systems. Niemiec and Kravaris [10] proposed a systematic procedure for the construction of statically equivalent outputs with prescribed transmission zeros. Kanter et al. [11] developed nonlinear control laws for input constrained, multiple-input, multiple-output, stable processes. They addressed the nonlinear control of the processes by exploiting the connections between model- predictive control and input-output linearization. Kravaris et al. [12] presented a systematic method arbitrarily assigning the zero dynamics of a nonlinear system by constructing the requisite synthetic output maps. The method requires solving a system of first-order, nonlinear, singular PDEs. Mickle et al. [13] developed a tracking controller for unstable, nonlinear processes by using trajectory linearization. Tomlin and Sastry [14] derived tracking control laws for non-linear systems with both fast and slow, possibly unstable, zero dynamics. Vander Scaft [15] developed a nonlinear state feedback H_∞ optimal controller. Devasia et al. [16] and Devasia [17] introduced an inversion procedure for nonlinear systems that constructs a bounded input trajectory in the pre-image of a desired output trajectory. Hunt and Meyer [18] showed that under appropriate assumptions the bounded solution of the partial differential equation of Isidori and Byrnes [19] for each trajectory of an ecosystem must be given by an integral representation formula of Devasia et al. [17] Chen and Paden [20] studied the stable inversion of nonlinear systems.

The present work uses the same continuous-time, model predictive control framework. Conceptually as the amount of nonlinearity is very less so the model does not require any linearization of the process equations. Here the nonlinear terms are considered as the strain rate in the convective heat transfer correlation. The controller introduced here is obtained by requesting desired responses for the state variables. The nonlinear state feedback is derived by minimizing a function norm of the deviations of the controlled outputs from linear reference trajectories with orders equal to the (output) relative orders. In the part of the work described in this chapter, we come up through a systematic evaluation with a process model that provides the starting point of evolving a model-based control. At the same time, we analyze the performance of a model-based nonlinear control with the applications of an efficient pole-placement technique. The focal point of the present analysis is a detailed modeling of the heat transfer and applications of efficient techniques (pole-placement) to control the state variables influencing the process. Moreover this chapter presents a nonlinear control method that is applicable to stable and unstable processes. The closed loop stability is ensured by forcing all the process state variables to follow their corresponding reference trajectories. The proposed control system includes a nonlinear state feedback and pole placement technique in both the forms of fixed poles and variable poles.

1.1. Pole placement techniques

“Pole placement” is a conventional design methodology for linear time-invariant control systems. The method is based on the fact that several performance specifications can be met by using dynamic output feedback to adequately placed closed loop poles in the complex plane. . An extension of the classical pole placement problem is the regional pole-placement problem, in which the objective is to place closed-loop poles in a suitable region of the complex plane. In many real cases, the model uncertainty reflects on the parameters of the plant, which has motivated extensive research efforts in parametric robust theory [21, 22, 23]. Pole placement methods can be divided mainly into two categories such as output feedback and state feedback methods. Usually dominant pole design techniques are employed as it is not always possible to assign all the poles of a system and thus one has to relax the design requirements and consider a partial pole placement that will give satisfactory overall performance. The pole placement problem is much easier to handle when the design is done in the state space domain and many techniques exist for this particular case. In this work, the plant is represented by proper state-space equation. Robust pole placement problem is formulated as the problem of robustly assigned poles in a region in two different ways i.e., by assuming some appropriate fixed poles and variable poles in terms of the Eigen values of the ($\mathbf{A} - \mathbf{BK}$) matrix. Where ‘ \mathbf{K} ’ is the gain. Similar conditions hold in [24] and [25].

The main focus of this work was to present a more efficient way of approaching Ackerman's formula to find the gain matrix so as to achieve the Eigen poles. Apart from this, the technique presented here was utilized towards significant progress for the case of nonlinear Single-input and multiple-output (SIMO) systems. . As only the temperatures of the plate are controlled, typically the plate may remain insignificantly cool or the desired cooling temperature trajectory is followed backwards. That is why we focus on state tracking instead of output tracking. To facilitate the design of controller, we first investigate the pole placement problem for SIMO systems using dynamic gains, when all the state variables are available for feedback. Both the local and global pole placement control has been accomplished in the MATLAB and SIMULINK environments. In both the cases the results are shown in the comparative mode between the demand and controlled states.

II. THE DESIGN PROBLEM

In this work, we consider the modeling to evolve a dynamic model-based control of the air jet impingement cooling of steel plate. Here we propose the design procedure based on state space pole placement techniques for the state feedback control design of jet impingement cooling system. The design of State-Space models is not different from that of transfer functions in that the differential equations describing the system dynamics are written first. For state-Space models, models, instead, the equations are arranged into a set of first order differential equations in terms of selected state variables, and the outputs are expressed these same state variables, Because the elimination of variables between equations is not an inherent part of this process, state models can be easier to obtain. The State models are directly derived from the original system equations. The standard form of the State-Space model equation is:

$$\dot{\mathbf{X}}(\mathbf{t}) = \hat{\mathbf{A}}\mathbf{X}(\mathbf{t}) + \hat{\mathbf{B}}\bar{\mathbf{a}}(\mathbf{t}) \text{ (State Equation)}$$

Here 'x' is the State Vector, the vector of state variables, \bar{a} is the control variable and 'A' is the system coefficient matrix. The matrices $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ are controllable. A fixed dynamic controller is designed i.e.,

$$\bar{\mathbf{a}} = \hat{\mathbf{B}}^{-1}[\hat{\mathbf{A}}\bar{\mathbf{X}} - \dot{\bar{\mathbf{X}}}]$$

The pole placement design of 'B' in the form of full state feedback becomes equivalent to design 'K' where,

$$\mathbf{a}_e = -\mathbf{K}\mathbf{X}_e \text{ (Controller or Control Law)}$$

(1)

For stability all the Eigen values of 'A' must be such that the real part is negative. Let us assume,

$$\mathbf{Z} = \mathbf{T}\mathbf{X} \tag{2}$$

The equation becomes,

$$\begin{aligned} \mathbf{T}\dot{\mathbf{X}} &= \mathbf{T}\mathbf{A}\mathbf{X} + \mathbf{T}\mathbf{B}\bar{\mathbf{a}} \\ \dot{\mathbf{Z}} &= \mathbf{T}\mathbf{A}\mathbf{T}^{-1}\mathbf{Z} + \mathbf{T}\mathbf{B}\bar{\mathbf{a}} \quad (\because \mathbf{X} = \mathbf{T}^{-1}\mathbf{Z}) \\ \dot{\mathbf{Z}} &= \bar{\mathbf{A}}\mathbf{Z} + \bar{\mathbf{B}}\bar{\mathbf{a}}; \end{aligned} \tag{3}$$

Where $\bar{\mathbf{A}} = \mathbf{T}\mathbf{A}\mathbf{T}^{-1}$ and $\bar{\mathbf{B}} = \mathbf{T}\mathbf{B}$

Let us consider the Eigen values of equation (3) be $\bar{\lambda}$,

$$\text{Then, } |\bar{\mathbf{A}} - \bar{\lambda}\mathbf{I}| = 0$$

$$|\mathbf{A} - \bar{\lambda}\mathbf{I}| = 0 \Rightarrow \lambda = \bar{\lambda}$$

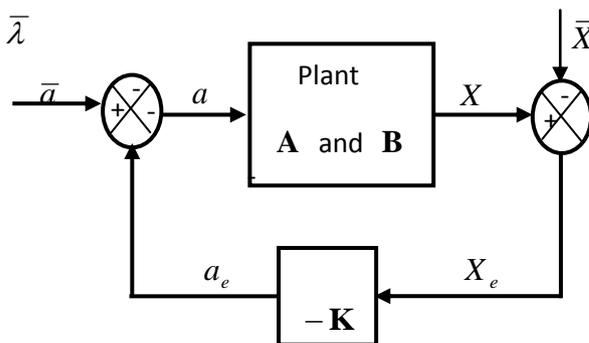


Figure 1: State Feedback Configuration

Since, $\bar{\lambda} = \lambda$, the transformation 'T' does not alter the system stability. The above diagram represents the corresponding control structure that evolves in the form of a feedback structure. In contrast to traditional feedback structure, **K** varies with time. Then feeding this input $\bar{\mathbf{a}}(\mathbf{t})$ back into the system, we obtained,

$$\dot{\mathbf{X}}(\mathbf{t}) = (\mathbf{A} - \mathbf{BK})\mathbf{X}(\mathbf{t}) \tag{4}$$

III. FULL STATE FEEDBACK DESIGN

There are several motivations for formulating the design procedure in state space. One motivation is a clear exposition of the controller structure. Another is that the controller structure points possibilities of reducing the order of the compensators using state space technique [26]. A third motivation is that the state space formulation may accommodate design procedures for systems in which the number of inputs or outputs is less than the number of operating conditions.

Let us consider the system P has $r > 1$ operating conditions with models $P_1(S), P_2(S), \dots, P_r(S)$. The state space representations of the models $P_i(S), i = 1, 2 \dots r$, are:

$$\dot{\mathbf{X}} = \mathbf{A}_i\mathbf{X} + \mathbf{B}_i\mathbf{a} \tag{5}$$

Where the state $\mathbf{X} \in \mathbf{R}^n$ and the input $\mathbf{a} \in \mathbf{R}^m$. The pairs $(\mathbf{A}_i, \mathbf{B}_i), i = 1, 2 \dots r$ are controllable. Let us design a fixed static controller $\mathbf{a} = -\mathbf{KX}$ such that the poles of the closed loop system are $\lambda(\mathbf{A}_i - \mathbf{B}_i\mathbf{K}) = \lambda_i, i = 1 \dots r$, where σ_i are the desired Eigen values for the i^{th} operating conditions. For full state to have a solution, it is necessary that the pairs $(\mathbf{A}_i, \mathbf{B}_i), i = 1 \dots r$, be controllable. Closely related to full state problem is the design of a fixed gain matrix **K** to stabilize all the operating conditions. The solution to problem exists a **K** such that:

$$\det(\mathbf{sI} - (\mathbf{A}_i - \mathbf{B}_i\mathbf{K})) = \mathbf{p}_i(\mathbf{s}), i = 1, \dots, r. \tag{6}$$

The nonlinear pole-placement equation (6) can be considered as Bass-Gura pole-placement formula [27], extended to multiple operating conditions.

IV. NUMERICAL METHODS FOR THE SINGLE POLE PLACEMENT PROBLEM

There are several methods for the pole placement problem some of the well known theoretical formulae, such as the Ackermann's formula, Bass-Gura formula etc.. The primary reason is that, these methods are based on transformation of the controllable pair $(\mathbf{A}_i, \mathbf{B}_i)$ to the controller companion form.

The computationally viable methods for pole placement are based on transformation of the pair $(\mathbf{A}_i, \mathbf{B}_i)$ to the controller Heisenberg form or the matrix **A** to the real Schur form (RSF), which can be achieved using orthogonal transformations. The pole placement design in the present study is accomplished in two ways, that is

- i) Assigning the fixed poles arbitrarily in the complex plane.
- ii) Assigning the variable poles in the complex plane.

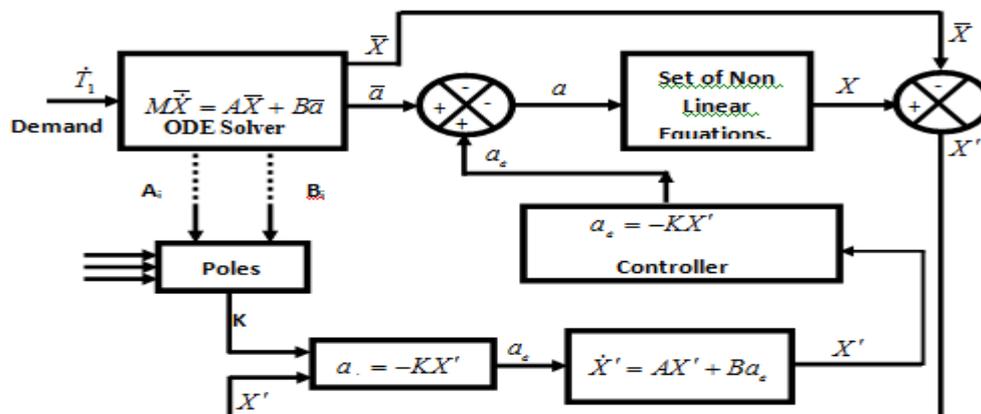


Figure 2: Block Diagram Showing Pole-Placement Control Model for Jet Impingement Cooling of Steel Plate (Temperature tracking Control).

V. STATE FEEDBACK CONTROL BY FIXED AND VARIABLE POLE PLACEMENT DESIGN

As a first attempt, a non-linear model-referenced state-feedback tracking control by fixed (local) pole placement for a jet impingement heat transfer system was implemented. The closed-loop poles were adjusted arbitrarily at each of the time intervals based on the open-loop poles of the system. The arbitrary pole assignment was performed in the SIMULINK environment. The poles were selected on the basis of trial and error method, for different simulation times. For each time segment the controlled output response and the reference output track were identified and compared. The results obtained in the SIMULINK environment for different time segments and time steps adopting the fixed pole placement control of the temperature track are discussed in detail in the later sections. After picking the feedback gain matrix \mathbf{K} to get the dynamics to have some nice properties (i.e., Stabilize A) $\lambda_i(\mathbf{A}) \rightarrow \lambda_i(\mathbf{A} - \mathbf{BK})$, the selected poles were placed in the complex plane. Recalling that the characteristic polynomial for this closed-loop system is the determinant of $(s\mathbf{I} - (\mathbf{A} - \mathbf{BK}))$. Since the matrices \mathbf{A} and $\mathbf{B}*\mathbf{K}$ are both 3 by 3 matrices, there were 3 poles for the system. By using full state feedback we placed the poles at different places arbitrarily. The control matrix \mathbf{K} , which gave the desired poles, was found by using the MATLAB function **PLACE**. For SIMO systems, **PLACE** use the extra degrees of freedom to find a robust solution for \mathbf{K} i.e., minimizes sensitivity of closed loop poles to perturbations in \mathbf{A} and \mathbf{B} .

Then, the pole placement design procedures have been executed for the state feedback design of systems with single operating condition. To facilitate the design, poles varying with time were introduced. The same controller using strictly the state space technique was used for simulation i.e., $\dot{\mathbf{a}} = -\mathbf{KX}$. This approach provided a clear exposition of the controller structures. The closed-loop poles were adjusted at each of the time intervals based on the open-loop poles of the system. The feedback gain was developed based on the plant models and the corresponding set of closed-loop poles. The response of the non-linear plant to this controller was presented for the full state feedback. In order to target a steady state of the system the Eigen values of $(\mathbf{A} - \mathbf{BK})$ should be such that their real parts are negative. They were chosen in the vicinity of the open-loop poles of the system. If any of the open-loop poles have positive real part, the real part of the corresponding closed-loop poles were chosen to be zero. The time variant Eigen values of the matrix $(\mathbf{A} - \mathbf{BK})$ were computed then computed.

VI. RESULTS AND DISCUSSIONS

The main focus of this work was to evolve a model based closed loop control analysis with the application of an efficient pole placement technique. On the way to establish the necessary control model, we captured an important jet impingement cooling of a steel plate. The simulations were carried out in the MATLAB and SIMULINK environment. Different sets of results were generated for different set of poles. The results discussed below are both for fixed closed loop pole placement and variable pole placement problems. In the closed loop control phenomena, the important task was evaluation of elements of the feedback gain matrix (\mathbf{K}).

6.1 Closed- loop control results for fixed pole placement technique

is observed that with the application of fixed poles such as two complex conjugate poles and one real pole, where all the real poles are negative i.e., poles are considered more in the left half of the complex plane, the cooling rate controls upto -1 K/s and the nature of controlled temperature track is almost coincides over the reference temperature track as shown in the figures 3 and 4. The poles selected initially are $-0.2 \pm 1i$ (Complex Conjugate) and -100 (Real). Several trials were carried out by changing the pole locations and plots were generated for the temperature track on the top surface of the steel plate to realize the controlled cooling rate.

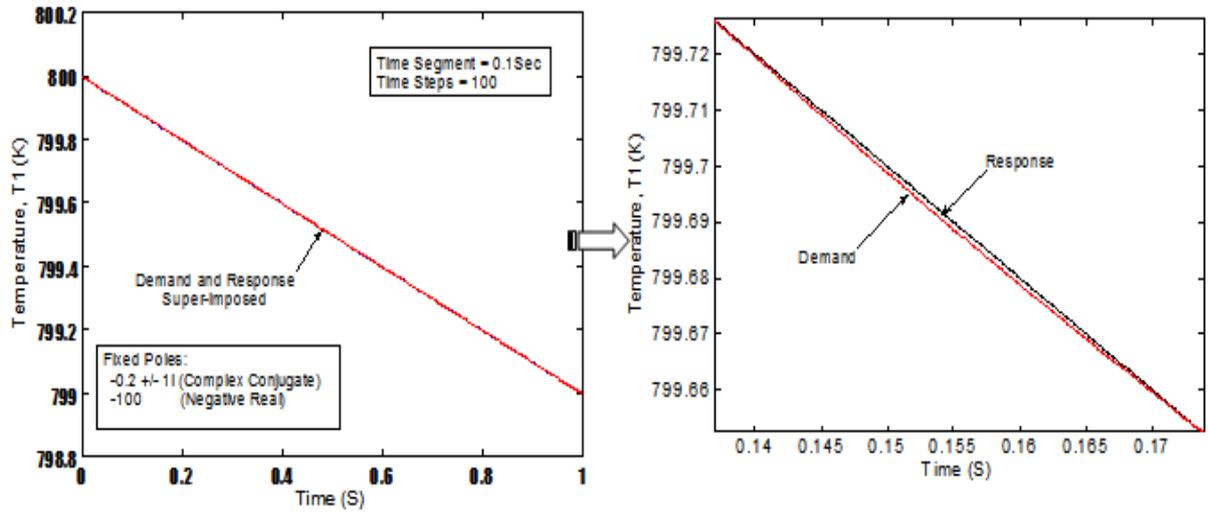


Figure 3: Effect of linearization time interval on tracking control of Plate temperature by assigning a fixed set of

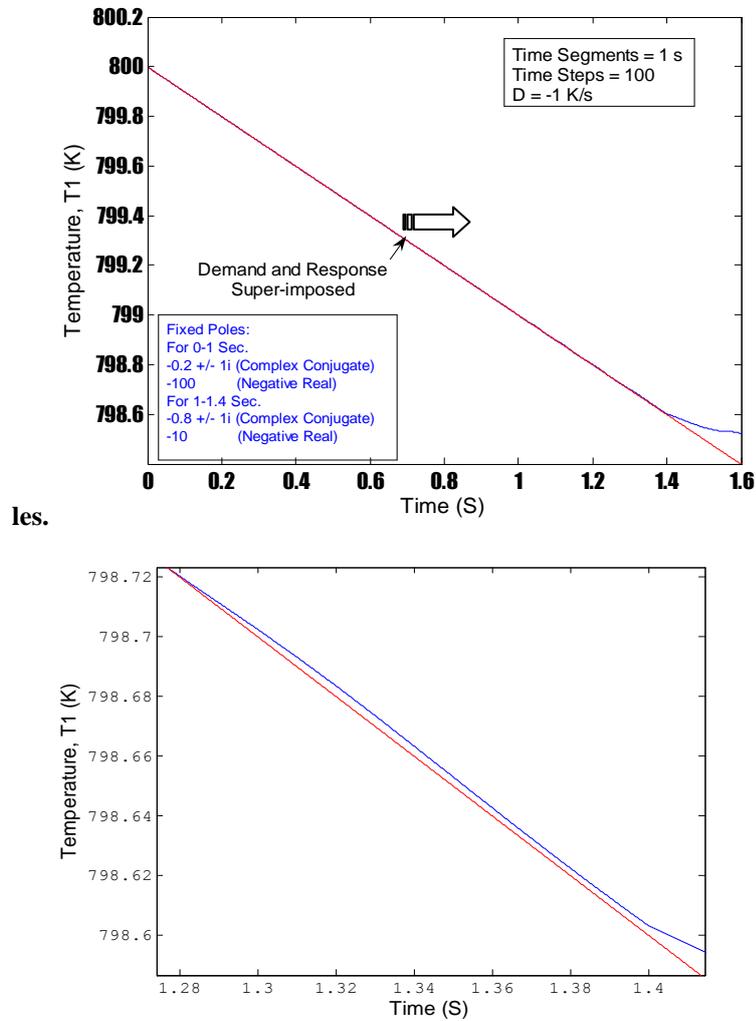


Figure 4: Effect of linearization time interval on tracking control of Plate temperature by assigning a fixed set of poles at Time Steps of 100 and Time segment of 1 second.

6.2 Closed- Loop control results for variable pole placement technique

This technique applies the Eigen values of the matrix (A) where, the matrix A is a 3 by 3 matrix, so there will be 3 poles for the system. Using the MATLAB function *eig*, the Eigen values of the (A) matrix are evaluated to evolve the feedback gain matrix (K). The problem was solved as a closed loop feedback control model using time dependent poles. For the positive values of the poles, the stability of the system was found marginal. Thus, at each linearized time segment the appropriate poles were assigned and the corresponding gain matrix was computed. The total computation has been performed by the help of SIMULINK tools. The effects of linearization time segment on tracking control of Plate temperature by assigning variable poles at different time locations are presented in the figures 5, 6, and 7 respectively. The only peculiarity found in this case was that, the response was oscillatory in nature. These results clearly predict that, one of the dominant poles tend to more negative side, while other two become less effective.

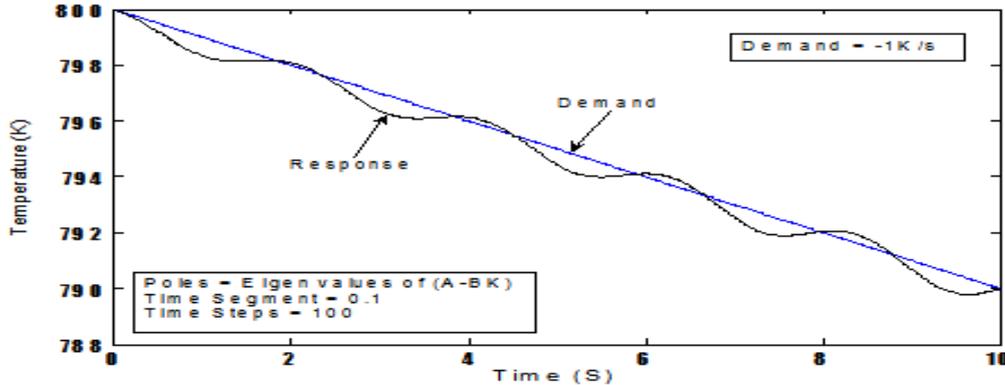


Figure 5: Effect of Linearization Time Interval on Tracking Control of Plate Temperature by Assigning a Set of Variable Poles at Different Time Locations at Time Steps of 100 and Time Segment of 0.1 second.

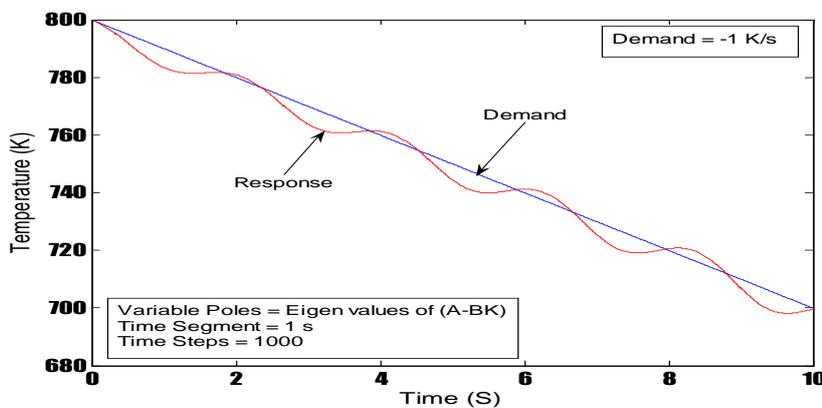


Figure 6: Effect of Linearization Time Interval on Tracking Control of Plate Temperature by Assigning a Set of Variable Poles at Different Time Locations at Time Steps of 1000 and Time Segment of 1 second.

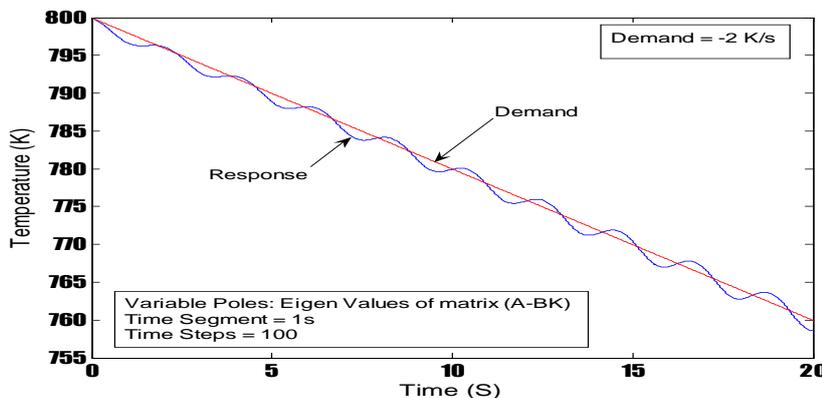


Figure 7: Effect of Linearization Time Interval on Tracking Control of Plate Temperature by Assigning a Set of Variable Poles at Different Time Locations at Time Steps of 100 and Time Segment of 1 second.

VII. CONCLUSION

A model based simulation and control of the temperature track in jet impingement cooling process of a steel plate has been developed that incorporates the implementation of an efficient pole placement technique in two ways i.e., one by the local poles and other by the global poles. The nonlinear control model developed here can be used to evaluate the performance in terms of physical properties and metallurgical aspects. The conclusions being obtained from the work are:

1. Pole placement design procedures have been proposed for the state feedback design of systems with single input condition. To facilitate the design, the concepts of coefficient matrices (A and B) and local poles are introduced.
2. It is seen that the response of the non-linear system is sensitive to linearization time interval. Better control is implemented by increasing the frequency of adjustment of the closed-loop poles.
3. In the present algorithm, the mean solution is utilized at discrete intervals (Piece-wise linearized time intervals) of time for evaluation of the coefficients.

It is observed from both the fixed pole placement and variable pole placement that, in case of fixed pole placement, the simulation time is limited upto 1.4 seconds. Within this simulation time, the demand temperature track and the response are coinciding to each other with very negligible error. While, in case of variable pole placement, the simulation time is large for the same linearized time intervals and percentage of error is significant compared to the first case.

REFERENCES

- [1]. D.Q. Mayne, J.B. Rawlings, C.V. Rao, P.O.M. Scokaert. constrained model predictive control: Stability and Optimality. *Automatica* 2000, 36 (6), 789.
- [2]. F. Allgower, A. Zheng. *Nonlinear Model Predictive Control; Progress in Systems and Control Theory Series*; Birkhauser Verlag; Basel 2000; Vol. 26.
- [3]. S.L. de Oliveira, M.V. Kothare, M. Morari. Contractive model predictive control for constrained nonlinear systems. *IEEE Trans. Autom. Control* 2000, 45, 1053.
- [4]. C. Kravaris, P. Daoutidis, R.A. Wright. Output feedback control of no minimum-phase nonlinear processes. *Chem. Engg. Sc.*, 1994, 49 (13), 2107.
- [5]. A. Isidori. *Nonlinear control systems*; Springer- Verlag; New York, 1995.
- [6]. N.H. El-Farra, P.D. Christofides. Integrating Robustness, Optimality and constraints in control of Non linear processes. *Chem. Engg. Sci.* 2001, 56, 1841.
- [7]. N.H. El-Farra, P.D. Christofides. Bounded Robust control of constrained nonlinear processes. *Chem. Engg. Sc.* 2003, 58, 3025.
- [8]. S. Djuljevic, N. Kazantzis. A New Lyapunov design approach for nonlinear systems based on Zubov's method. *Automatica* 2002, 38, 1999.
- [9]. C. Kravaris, P. Daotidis. Nonlinear State feedback control of second order non-minimum phase nonlinear systems. *Comput. Chem. Eng.* 1990, 40 (4/5), 439.
- [10]. M. Niemiec, C. Kravaris. Nonlinear model-state feedback control for non-minimum -phase processes. *Automatica* 2003, 39, 1295.
- [11]. J.M. Kanter, M. Soroush, W.D. Seider. Nonlinear controller design for input-constrained, multivariable processes. *Ind. Eng. Chem. Res.* 2002, 41, 3735.
- [12]. C. Kravaris, M. Niemiec, N. Kazantzis. Singular PDEs and the assignment of zero dynamics in nonlinear systems. *Syst. Control Lett.* 2004, 51, 67.
- [13]. M.C. Mickle, R. Huang, J.J. Zhu. Unstable, non-minimum phase, nonlinear tracking by trajectory linearization control. *Proceedings of the IEEE International Conference on Control applications, Taipei, Taiwan, 2004*; p812.
- [14]. C.T. Tomlin, S.S. Sastry. Bounded tracking for non-minimum phase nonlinear systems with fast zero dynamics. *Int. J. Control* 1997, 68, 819.
- [15]. A.J. Van der Schaft. L_2 - Gain analysis of nonlinear systems and nonlinear state feedback H_∞ control. *IEEE Trans. Autom. Control* 1992, 37, 770.
- [16]. S. Devasia, D. Chen, B. Paden. Nonlinear inversion-based output tracking. *IEEE Trans. Autom. Control* 1996, 41, 930.
- [17]. S. Devasia. Approximated stable inversion for nonlinear systems with non-hyperbolic internal dynamics. *IEEE Trans. Autom. Control* 1999, 44, 1419.
- [18]. L.R. Hunt, G. Meyer. Stable inversion for nonlinear systems. *Automatica* 1997, 33, 1549.
- [19]. A. Isidori, C.I. Byrnes. Output regulation of nonlinear systems. *IEEE Trans. Autom. Control* 1990, 35, 131.
- [20]. D.G. Chen, B. Paden. Stable inversion of nonlinear non-minimum phase systems. *Int. J. Control* 1996, 64, 81.
- [21]. J. Ackermann. (1993). *Robust Control: Systems with Uncertain physical parameters*. Springer-Verlag, New York, NY.
- [22]. B.R. Barmish. (1994). *New tools for robustness of linear systems*. Macmillan Publishing Co., New York, NY.
- [23]. S.P. Bhattacharya, H. Chapellat, L.H. Keel. (1995). *Robust Control- The parametric approach*. Prentice Hall publishing Co, Upper Saddle River, NJ.
- [24]. Y.C. Soh, R.J. Evans, I. Petersen, R.E. Betz. Robust Pole assignment. *Automatica*, 1987, 23, pages 601-610.
- [25]. L.H. Keel, S.P. Bhattacharya. A linear programming approach to controller design. *Proceedings of the 36th Conference on Decision and control.* 1997, pages 2139-2148, San Diego, CA, USA.
- [26]. T. Kailath, *Linear systems*, Anglewood Cliffs, Prentice-Hall, 1980.
- [27]. R. W. Bass, I. Gura, "High order design via State space considerations", *Proc. 1965 JACC*, pp. 311-318.
- [28]. Chow, J.H. A pole placement design approach for systems with multiple operating conditions, *Proc. 27th Conf. On Decision and Control, Austin, Texas.* 1988, pages 1272-1277.

Remote Control System through Mobile and DTMF

Abdiweli Abdillahi Soufi¹, Abdirasoul Jabar Alzubaidi²

¹ Information Technology College, Open University of Sudan.

² Electronics Engineering School, Sudan University of Science and Technology

ABSTRACT:

This paper proposes a control system which enables controlling remotely through mobile phone and DTMF decoder. The system will control a cutter machine in foam processing factory. The proposed system will take remote control process over mobile network. In this system a mobile phone sends controlling signal from remote location. 3.75G internet modem is used to receive the signal from the mobile and transmits the signal to DTMF decoder through computer headphone, DTMF 8870 decoder is used to decode the tone generated by the buttons of the mobile keypad. PC computer hosts software for controlling cutter machine, latching IC in the system is for signal buffering and Darlington IC for signal amplifying. The proposed system provides solution for industrial sector to access and control devices and machines remotely.

KEYWORDS: Controlling signal, Cutter Machine, DTMF Decoder, Internet Modem, Latching IC, Mobile Phone, PC computer, and Signal tone.

I. INTRODUCTION

Remote Control System is a control system that allows accessing and managing an electrical or electronic device remotely [2]. Remote control systems are used in different fields like home management, factory automation, space exploration, robot operating, and other critical fields [1,5]. There are many advantages of using remote control systems; remote control systems are great time and effort savers, also they provide flexibility, security, independence of location or geographical distance.

This paper proposes a method to control remote machine using mobile and DTMF decoder. The proposed system will take remote control process over mobile network to control a cutter machine, without regarding the phone model and mobile phone carrier. Using mobile networks in remote control system has following characteristics: Reliable and resilient, immune to interference from other radio sources, data and internet oriented, broadband capability, and wider coverage [7]. This system uses a mobile phone and internet modem. PC computer is a central unit for the system which hosts software program to manage controlling process. Many proposed systems for remote controlling uses two mobiles to send and receive controlling signal from the user to the side where controlled device is allocated. In this system a mobile phone and 3.75G internet modem are used, the modem can receive call from the mobile, since the modem is used primarily for internet usage using controlling purpose can be useful. The proposed system will control cutting machine in foam processing factory. The machine cuts foam blocks by rotating the blocks into different predesigned sizes. The system can carry following commands to control cutting machine; Turn-on/turn-off machine, change foam size to predesigned shapes (designs one, design two, etc.), Increase machine speed, decrease machine speed.

II. SYSTEM COMPONENTS

DTMF Decoder

DTMF is short for Dual Tone Multi Frequency. It is a generic communication term for touch tone (a Registered Trademark of AT&T). The tones produced when dialing on the keypad on the phone could be used to represent the digits, and a separate tone is used for each digit. Pressing any key generate unique tone which consists of two different frequencies one each of higher and lower frequency range. The resultant tone is convolution of two frequencies [1,5,6]. Figure 1 shows mobile keypad frequencies and table1 also shows tone frequency associated with a particular key.

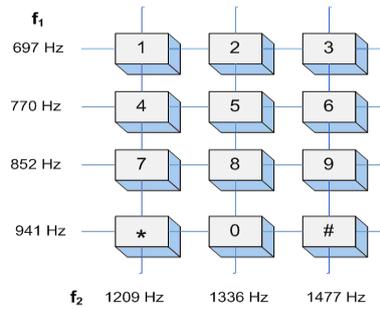


Figure 1 Phone keypad for DTMF generation [10]

Table1 Tone frequency associated with a particular key resulting

Button	Low Frequency	High Frequency	Key Frequency
1	697	1209	1906
2	697	1337	2034
3	697	1477	2174
4	770	1209	1979
5	770	1337	2107
6	770	1477	2247
7	852	1209	2061
8	852	1337	2189
9	852	1477	2329
0	941	1209	2150
*	941	1337	2278
#	941	1477	2418

Each of these tones is composed of two pure sine waves of the low and high frequencies superimposed on each other. These two frequencies explicitly represent one of the digits on the telephone keypad. Thus generated signal can be expressed mathematically as follows:

$$f(t) = AH \sin(2\pi fH t) + AL \sin(2\pi fL t) \quad (1)$$

Where AH, AL are the amplitudes & fH, fL are the frequencies of high & low frequency range. Properties of DTMF tone frequencies are:

- No frequency is an integer multiple of another
- The difference between any two frequencies does not equal any of the frequencies
- The sum of any two frequencies does not equal any of the frequencies

1. Sudani One 3.75G Internet modem:

Sudani One 3.75G Is a USB modem for wireless internet connections, based on 3.75G (HSDPA/HSUPA) technology with speeds up to 7.2 Mb/s to provide convenient and flexible Internet access, whether the user machine is PC computer , laptop, Mobile or any other device[8].



Figure 2 Sudani One Modem[8]

2. PC Computer:

PC computer hosts developed software using C++ programming language to control cutting machine. The PC computer is connected with 8870 DTMF decoder via parallel port inputs. The software dictates the processor to handle controlling process. A corresponding signal is then sent via the output pins of the parallel port to the HD74LS373 latch IC.

3. HD74LS373 Latching IC:

The HD74LS373 is eight bit register IO mapped used as a buffer which stores signals for future use. Different types of latches are available HD74LS373 octal D-type transparent latch will be used in this system. This type of latch is suitable for driving high capacitive and impedance loads[12].

4. ULN 2001A Darlington IC:

The ULN2803A is a high-voltage, high-current Darlington transistor array. The device consists of eight NPN Darlington pairs that feature high-voltage outputs with common-cathode clamp diodes for switching inductive loads. The collector-current rating of each Darlington pair is 500 mA. The Darlington pairs may be connected in parallel for higher current capability[11].

5. Cutting Machine:

Cutting machine is an electrical machine used to cut raw blocks of expanded foam into predesigned shapes. The machine uses a type of tool known as a hot-wire foam cutter. The predesigned shapes are both manual and fully automatic designs. The applied cutting machine cuts foam blocks by rotating the block.



Figure 3 cutter Machine

III. METHODOLOGY

The main goal of the proposed system is to send controlling signal remotely from mobile phone to controlled machine using mobile network. The whole system can be divided into following stages:

a. Mobile Phone Stage:

A mobile phone is used to send controlling signal. The mobile user dials SIM card number in the Sudani One modem. After initiating the call, pressing particular button of mobile keyboard produces specific DTMF tone.

b. Internet Modem Stage:

Sudani One modem is used in this system but any other internet modem contains SIM card can be used. Sudani One modem is connected with PC computer through USB port. The modem is auto-receiving mode. The mobile and the modem are connected via mobile network by receiving call from the mobile. After receiving the control signal the modem transmits the signal to the PC through.

c. 8870 DTMF Decoder Stage:

The signal is fed to 8870 DTMF decoder chip through headphone connected with the PC computer. The DTMF decoder will give corresponding BCD value of tone as shown in table2. Each key press at transmitter end reflects as a BCD value of Q1, Q2, Q3, and Q4. The outputs of the DTMF decoder are sent to the PC through parallel port [5].

Table2 8870 DTMF output truth table

Button	Low Frequency	High Frequency	TO E	Q4	Q3	Q2	Q1
1	697	1209	1	0	0	0	1
2	697	1336	1	0	0	1	0
3	697	1477	1	0	0	1	1
4	770	1209	1	0	1	0	0
5	770	1336	1	0	1	0	1
6	770	1477	1	0	1	1	0
7	852	1209	1	0	1	1	1
8	852	1336	1	1	0	0	0
9	852	1477	1	1	0	0	1
0	841	1209	1	1	0	1	0
*	841	1336	1	1	0	1	1
#	841	1477	1	1	1	0	0

0= LOGIC LOW, 1= LOGIC HIGH

d. PC Computer Stage:

The output signal from the 8870 DTMF decoder is fed to the computer via parallel port inputs. The software dictates the processor to make a sequential pulling for the parallel port control to check if any signal is available from input pins. The detected signal is handled by the Micro-processes according to the instructions in the software. A corresponding signal is then sent via the output pins of the parallel port to the HD74LS373 latch IC.

e. Latching Block Stage:

The HD74LS373 is eight bit is register IO mapped and has unique address that work with parallel port. The signal received from the parallel port of the PC computer must be latched to be available to the interface to be controlled till a further change in the signal occurs.

f. Darlington Pairs Block:

The outputs of the latch pins are fed to the input pins of the ULN 2001A Darlington pair chip, this chip is the ideal solution for the interface between the low level logic circuitry and multiple peripheral power loads.

IV. BLOCK DIAGRAM

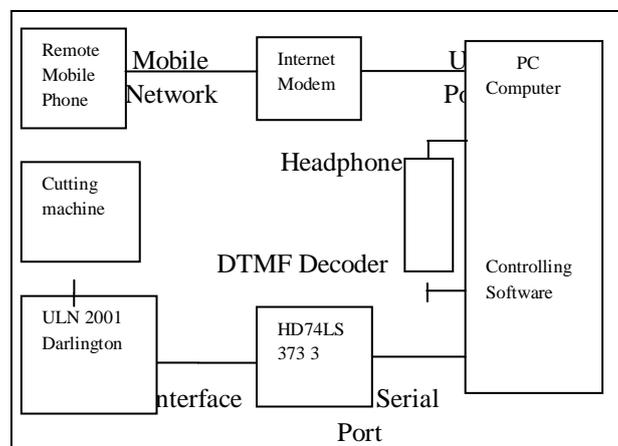


Figure 4 Block diagram

SYSTEM ALGORITHM AND FLOWCHART

Following Algorithm and flowchart are designed to control mixing machine remotely.

a) System Algorithm

1. Make Parallel port address (0x378) as outputport and Parallel port address (0x379) as inputport
2. Put outputport off (outputport=0x00)
3. Read input from inputport
4. Check if the input at inputport is 0001 then
Turn on cutter machine (output=0x01)
Go to step 4
Else go to step 6
5. Check if the input at inputport is 0010 then
Design One (output=0x02)
Go to step 4
Else go to step 7
6. Check if the input at inputport is 0011 then
Design two (output=0x04)
Go to step 4
Else go to step 8
7. Check if the input at inputport is 0100 then
Design three (output=0x06)
Go to step 4
Else go to step 9
8. Check if the input at inputport is 0101 then
Increase the speed (output=0x08)
Go to step 4
9. Else go to step 10
Check if the input at inputport is 0110 then
Decrease the speed (output= (0x010))
Go to step 4
Else go to step 11
Check if the input at inputport is 0111 then
Turn of the machine (output=0x00)
Go to step 4
Else go to step 12
If the user wants to continue
Go to step 4
Else Finish

b) System Flowchart

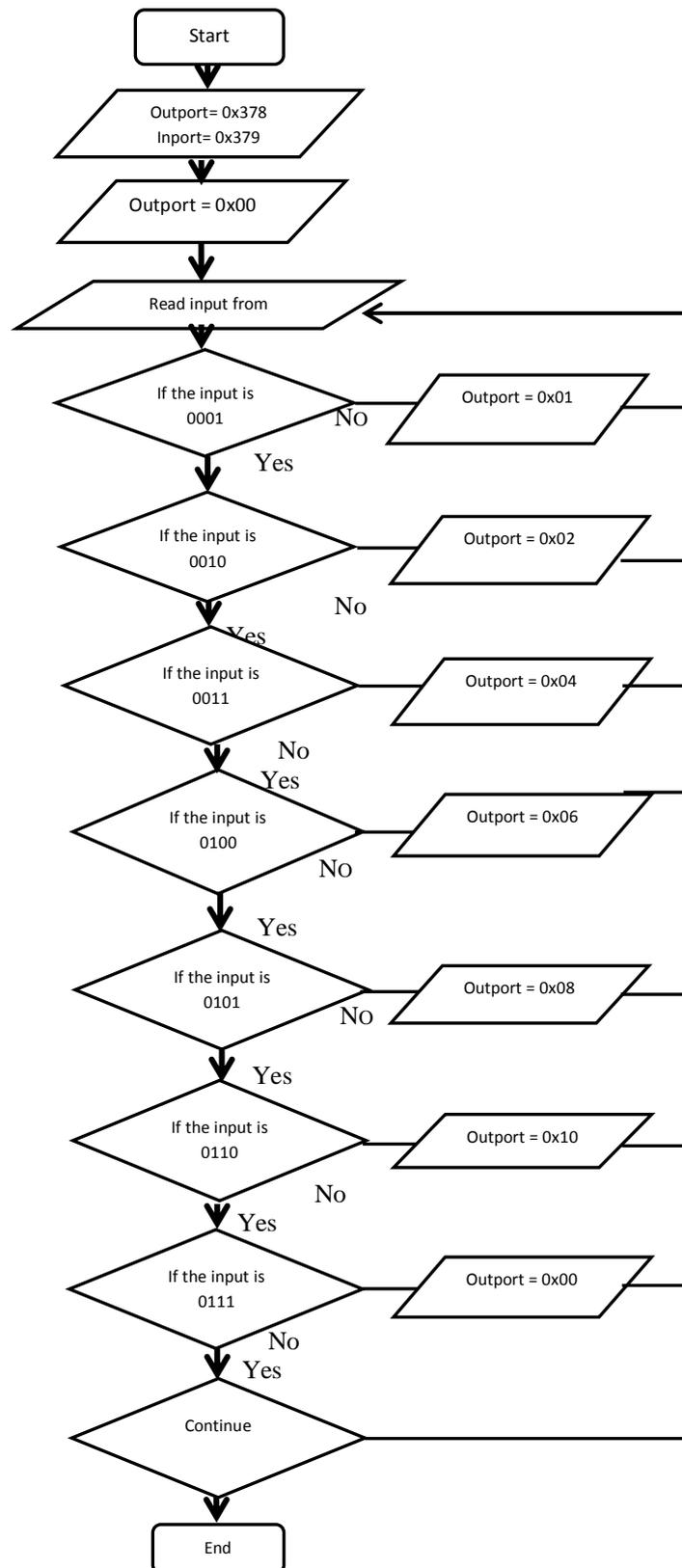


Figure 5 System flowchart

V. RESULTS

The controlling process for cutting machine was implemented successfully. The system was connected with the machine and executed all commands with unlimited number of times, following tables describes results of system implementation.

Table3 Implemented system Commands

Pressed key	Command	Result
1	Turn on	The machine is started
2	Design one	Changed to design one
3	Design two	Changed to design two
4	Design three	Changed to design three
5	Increase rotation speed	Increased
6	Decrease rotation speed	Decreased
7	Turn off	The machine is stopped

Table4 PC computer input and output results

Key Pressed	DTMF 8870 OUTPUT (BCD)	Input from Parallel Port				Output to Latching IC Pins (2,3,4,5,6,8,9)
		Pin 13	Pin 12	Pin 11	Pin 10	
1	0001	0	0	0	1	0x01
2	0010	0	0	1	0	0x02
3	0011	0	0	1	1	0x04
4	0100	0	1	0	0	0x06
5	0101	0	1	0	1	0x08
6	0110	0	1	1	0	0x010
7	0111	0	1	1	1	0x00

VI. CONCLUSION

This paper presented a method to control cutting machine using mobile phone and DTMF decoder. DTMF tone is generated by pressing the keypad buttons of the mobile which is connected to internet modem through mobile network. The popularity and availability of the mobile and mobile network makes this kind of control very useful and powerful. The main advantages of the proposed system are its reliability, low cost, and wide area coverage. Future works for this system can be following

- Adding SMS message to carry controlling commands as alternate way for DTMF tone.
- Upgrading the system to control more than one machine at same time.

REFERENCES

- [1] Roshan Ghosh. DTMF Based Controller for Efficiency Improvement of a PV Cell & Relay Operation Control Smart Home Systems, International Journal of Engineering Research and Applications. Vol. 2, Issue 3, May-Jun 2012, pp.2903-2911
- [2] Malik Sikandar, Hayat Khiyal, Aihab Khan, and Erum Shehzadi. SMS Based Wireless Home Appliance Control System (HACS) for Automation Appliances and Security. Journal of Informing Science and Information Technology, Volume 6 ,2009.
- [3] Kamrul Hassan, Raziul Islam Siddiqui, Md. Takdirul Islam, Nahid Alam Siddique, and Syed Mohammad Enam Uddin, GSM Based Automatic Motor Control and Protection System, International Journal of Advancements in Research & Technology, Volume 2, Issue2, Feb ruary-2013.
- [4] K.Arana, A.Sri Ramsagar, and G.Venkateswarlu, Mobile Operated Landrover Using Dtmf Decoder, International Journal of Modern Engineering Research, Vol.3, Issue.2, March-April. 2013 pp-898-902.
- [5] Tuljappa M Ladwa, Sanjay M Ladwa, R Sudharshan Kaarthik, Alok Ranjan Dhara and Nayan Dalei, Control of Remote Domestic System Using DTMF, ICICI-BME 2009 Bandung, Indonesia.
- [6] Er. Zatin Gupta, Payal Jain and Monika, A2Z Control System – DTMF Control System, Global Journal of Computer Science and Technology, Volume 10, Issue 11, October 2010.
- [7] http://www.smerobot.org/08_scientific_papers/papers/Calcagno_et-al_ISR-Robotik06.pdf
- [8] <http://www.sudani.sd/PublicOne/Content/Internet/HSDPA?language=en>
- [9] YUAN Xiao-chen, CHEANG Chak-Fong and Li Jian-Qing, Implementation of a Remote Control System for Smart Home.
- [10] <http://cnx.org/content/m19843/latest/>
- [11] www.dz863.com/datasheet-82174063-ULN2803A_Darlington-Transistor-Array/
- [12] <http://www.datasheetdir.com/NATIONAL-DM74LS373+Latches>

Seismic Response of Multi-Story Structure With Semi-Active Multiple Tuned Mass Friction Dampers

Alka Y. Pisal¹ and R. S. Jangid²

¹ Research Scholar , ² Professor

^{1,2}Department of Civil Engineering, Indian Institute of Technology Bombay, Powai, Mumbai - 400 076, India

ABSTRACT

A passive tuned mass friction damper (P-TMFD) has a pre-determined and fixed slip force at which it controls the response of the structure effectively and at any other values it loses its efficiency. To overcome this disadvantage, semi-active multiple tuned mass friction dampers (SA-MTMFDs) are proposed. The predictive control law is used for the semi-active system, so that it can produce continuous and smooth slip forces and eliminates the high frequency response of the structure which usually occurs in case of passive multiple tuned mass friction dampers (P-MTMFDs). Also, the effectiveness of SA-MTMFDs over P-MTMFDs is investigated. The governing differential equations of motion are solved numerically using the state-space method. The response of a five story shear type structure is investigated for four considered earthquake ground motions. The number of tuned mass friction damper (TMFD) units of SA-MTMFDs and P-MTMFDs are varied and the responses of the five story structure with SA-MTMFDs are compared with the responses of the same structure with the P-MTMFDs. For a fair comparison the displacement and acceleration response time histories of the structure with SA-MTMFDs and P-MTMFDs are compared with respect to their optimum controlling parameters. The result of numerical studies indicated that the SA-MTMFDs are more effective and has better performance level than the P-MTMFDs for same input seismic excitations.

KEYWORDS: P-MTMFDs, SA-MTMFDs, predictive control, numerical analysis, seismic excitation, optimum parameters and response reduction.

I. INTRODUCTION

The tuned mass damper (TMD) is a most popular and extensively used device to control vibrations in civil and mechanical engineering applications ranging from small rotating machinery to tall civil engineering structures. Similar to TMD, friction dampers (FD) were found to be very efficient, not only for rehabilitation and strengthening of existing structures but also for the design of structures to resist excessive vibrations (Colajanni and Papia, 1995; Qu et al., 2001; Mulla and Belev, 2002; Pasquin et al., 2004). In the past, some researchers had proposed the use of FD along with TMD. Ricciardelli and Vickery (1999) considered a SDOF system to which a TMD with linear stiffness and dry friction damping was attached. The system was analyzed for harmonic excitation and design criteria for friction TMD system were proposed. Lee et al. (2005) performed a feasibility study of tunable friction damper. Gewei and Basu (2010) used harmonic and static linearization solutions to analyze dynamic characteristics of SDOF system with friction tuned mass damper. P-TMFD is having advantage that it can behave either as a FD when it is in slip-state and as an added mass when it is in stick state. On the other hand, the main disadvantage of a single P-TMFD is its sensitivity of the effectiveness to the error in the natural frequency of the structure. If the design parameters of the TMD are chosen wrongly, it may accelerate the vibration of the system instead of attenuating it. To overcome this difficulty, many researchers proposed the use of more than one TMD with different dynamic characteristics, (Xu and Igusa, 1992; Joshi and Jangid, 1997). They proved that MTMDs are more effective than single TMD. The another disadvantage of P-TMFDs is that it has a pre-determined and a fixed value of slip force at which it reduces the response of the system to which it is attached, when it is in slip mode. At too small and too high value of slip force, the damper will not slip for the most of the harmonic and earthquake excitation duration and thus the capacity of P-TMFDs to reduce structural response may not be fully utilized. Also during an earthquake P-TMFD vibrate in two different modes (i.e. stick state and slip state), many times which results in high-frequency structural responses which are undesirable.

In order to improve the performance of such passive devices, the concept of semi-active control was emerged. The advantage of semi-active control system is that it is able to adjust its slip force by controlling its clamping force in real time with respect to the response of the structure during an excitation. Dowdell and Cherry (1994) and Kannan et al. (1995) were among the first researchers to study the response of structures with semi-active friction dampers. They adopted on-off and bang-bang control methods for their study. Inaudi (1997) proposed modulated homogeneous friction control algorithm which produces a slip force proportional to the prior local peak of the damper deformation. Akbay and Aktan (1995) proposed a control algorithm that determines the clamping force in next time step by one pre-specified increment of the current force at a fixed time step. Also, the literature review shows that the control performance of semi-active dampers fully depends on the applied control algorithm. There have been many studies on the development of the control law. From the review of the literature of these studies, it is clear that most of the developed algorithm either produces the discontinuous control forces or partially continuous slip forces. In both the cases the damper capacity may not be fully used. Recently, Lin et al. (2010) proposed SAF-TMD and investigated the effectiveness of SAF-TMD in protecting structures subjected to seismic forces using non-sticking law.

It is also observed that the semi-active control algorithms are developed specifically for TMD and for FD, but limited algorithms are developed for MTMFDs. In this study, the performance of SA-MTMFDs attached to a damped five story shear type structure is investigated for seismic ground excitations. The control algorithm developed by Lu (2004), known as predictive control is applied to SA-MTMFDs to get a continuous smooth slip force, so that it remains in its slip state during entire earthquake duration. The specific objectives of the study are summarized as (i) to formulate the equations of motion and obtained the response of Multi-story structure with SA-MTMFDs, under seismic excitations, (ii) to identify a appropriate parameter which controls the desired responses of the multi-story structure with SA-MTMFDs, (iii) to investigate the effect and optimum value of gain multiplier for the response reduction of the controlled multi-story structure and (iv) to investigate the effectiveness of SA-MTMFDs in response reduction under the earthquake excitations.

II. MODELING OF MDOF SYSTEM WITH SA-MTMFDs

The system configuration considered for the study consists of a primary system of five story shear structure attached with SA-MTMFDs with different dynamic characteristics as shown in Figure 1. The i^{th} floor of primary structure is characterized by mass m_i , stiffness k_i and fundamental frequency ω_s . The mass, stiffness and natural frequency of j^{th} SA-TMFD unit is shown as $m_{Td j}$, $k_{Td j}$ and $\omega_{Td j}$, respectively. The primary structure and each SA-TMFD unit are modeled as single degree of freedom system (SDOF) so that the total degrees of freedom of the system configuration considered for the study is $r+n$ where, r denotes the number of TMFD units and n denotes the number of degrees of primary structure. For the present study, the following assumptions are made:

- [1] The structural configuration of the primary system i.e. mass and stiffness of each floor are same. Also, the damping ratio for each mode of vibration is assumed to be constant.
- [2] Stiffness of each SA-TMFD unit is same.
- [3] The mass of each SA-TMFD unit is varying. By keeping the stiffness of each SA-TMFD constant and varying the mass, we vary the natural frequency of each SA-TMFD unit.
- [4] The natural frequencies of the SA-TMFDs are uniformly distributed around their average natural frequency. It is to be noted that SA-MTMFDs with indistinguishable dynamic characteristics are equivalent to a single SA-TMFD in which the natural frequency of the individual SA-MTMFD unit is same as that of the equivalent single SA-TMFD.
- [5] The SA-TMFD units apply variable friction force on the primary system, which can be controlled by varying the clamping force.

However, the mass and friction force is the sum of all the SA-MTMFDs masses and friction forces. Further, the system parameters used for the present study are described in detail below.

Let ω_T is the average frequency of all SA-MTMFDs, given as

$$\omega_T = \sum_{j=1}^r \frac{\omega_j}{r} \quad (1)$$

where, r is the total number of SA-MTMFDs. The natural frequency ω_j of the j^{th} SA-TMFD is expressed as

$$\omega_j = \omega_T \left[1 + \left(j - \frac{r+1}{2} \right) \right] \frac{\beta}{r-1} \tag{2}$$

where, β is the non-dimensional frequency spacing of the SA-MTMDs, given as

$$\beta = \frac{\omega_r - \omega_1}{\omega_T} \tag{3}$$

If $K_{Td j}$ is the constant stiffness of each j^{th} SA-TMFD, then the mass of the j^{th} SA-TMFD is expressed as

$$m_{Td j} = \frac{K_{Td j}}{\omega_j^2} \tag{4}$$

The ratio of the total SA-MTMFDs mass to the total mass of the main system is defined as the mass ratio and is expressed as

$$\mu = \frac{\sum_{j=1}^r m_{Td j}}{m_s} \tag{5}$$

where, m_s denotes the total mass of the main system.

The ratio of average frequency of the SA-MTMFDs to the fundamental frequency of main system is defined as tuning ratio, expressed as

$$f = \frac{\omega_T}{\omega_s} \tag{6}$$

III. EQUATIONS OF MOTION UNDER EARTHQUAKE EXCITATION

The governing equations of motion of multi-degree of freedom system (MDOF) with SA-MTMFDs when subjected to earthquake excitations are expressed as

$$M\ddot{X}_{(t)} + C\dot{X}_{(t)} + KX_{(t)} = E\ddot{x}_{g(t)} + BF_{d(t)} \tag{7}$$

where, the M , C and K denotes the mass, damping and stiffness of the configured system, considered for the study. $\ddot{X}_{(t)}$, $\dot{X}_{(t)}$ and $X_{(t)}$ are the relative acceleration, velocity and displacement vectors of the configured system relative to the ground. $\ddot{x}_{g(t)}$ represents vector of the ground acceleration and $F_{d(t)}$ represents the vector of the controllable friction forces provided by the SA-TMFDs. These matrices can be shown as:-

$$M = \begin{bmatrix} M_p & 0 \\ 0 & M_{Td} \end{bmatrix} \tag{8}$$

$$K = \begin{bmatrix} K_p + \sum_{j=1}^r K_{Td j} & -K_{Td1} & -K_{Td2} & -K_{Td3} & \dots & -K_{Td j} \\ -K_{Td1} & K_{Td1} & 0 & 0 & \dots & 0 \\ -K_{Td2} & 0 & K_{Td2} & 0 & \dots & 0 \\ -K_{Td3} & 0 & 0 & K_{Td3} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ -K_{Td j} & 0 & 0 & 0 & 0 & K_{Td j} \end{bmatrix} \tag{9}$$

$$C = \begin{bmatrix} C_p & 0 \\ 0 & 0 \end{bmatrix} \tag{10}$$

where, M_p , K_p and C_p represents the square matrices of dimensions $(n \times n)$, denotes the mass, damping and stiffness of primary five storey structure and n denotes the degrees of freedom of the primary structure. M_{Td} denote the square matrix of dimension $(r \times r)$, where r is the number of SA-TMFD units.

It is also to be noted that as the damping matrix of the primary structure is not known explicitly, it can be constructed using the Rayleigh's damping considering it proportional to mass and stiffness of the primary structure as,

$$C_p = a_0 M_p + a_1 K_p \tag{11}$$

where a_0 and a_1 are the coefficients which depends on the damping ratio of two vibration mode. For the considered primary structure, damping ratio is taken as 2% for all the modes of vibration and,

$$F_d = \begin{bmatrix} \sum_{j=1}^r F_{dj} & -F_{d1} & -F_{d2} & \dots & -F_{dr} \end{bmatrix} \tag{12}$$

$$X_{(t)} = \begin{Bmatrix} x_{p(t)} \\ x_{td(t)} \end{Bmatrix} \quad \text{and} \quad Z_{(t)} = \begin{Bmatrix} X_{(t)} \\ \dot{X}_{(t)} \end{Bmatrix} \tag{13}$$

Here, $x_{p(t)}$ represents the displacement of i^{th} floor of primary structure and $x_{td(t)}$ represents the displacement of j^{th} SA-TMFD unit of SA-MTMFDs respectively, relative to the ground. Also, the matrix E and B are placement matrices for the excitation force and friction force, respectively.

IV. SOLUTION OF EQUATIONS OF MOTION

Equation (7) can be formulated in dynamic state space as

$$\dot{Z}_{(t+1)} = AZ_{(t+1)} + E\ddot{x}_{g(t+1)} + BF_{d(t+1)} \tag{14}$$

where, vector $Z_{(t)}$ denotes the state of the system as shown in equation (13), $F_{d(t)}$ denotes the vector of controllable friction forces provided by the SA-TMFDs, $\ddot{x}_{g(t)}$ is the ground acceleration, A represents the system matrix that is composed of structural mass, stiffness and damping matrices. When the equation (15) is further discretized in the time domain assuming excitation force to be constant within any time interval, equation (15) can be converted into a discrete time form as mentioned by Meirovitch (1990).

$$Z_{(s+1)} = A_d Z_{(s)} + E_d \ddot{x}_{g(s)} + B_d F_{d(s)} \tag{15}$$

where, subscripts (s) and $(s + 1)$ denotes that the variables are evaluated at the $(s)^{th}$ and $(s + 1)^{th}$ time step.

$$B_d = A^{-1}(A_d - I)B \tag{16a}$$

$$E_d = A^{-1}(A_d - I)E \tag{16b}$$

Also, $A_d = e^{A\Delta t}$ denotes the discrete-time system matrix with Δt as the time interval.

Let y be a vector showing dampers displacements which are, $y = x_{td(j)} - x_5$, where $x_{td(j)}$ denotes the displacement of j^{th} SA-TMFD of SA-MTMFDs and x_5 denotes the displacement of the top i.e. fifth story. At any instant of time the relation between dampers displacements y and state of the structure Z may be written as

$$y_{(s)} = DZ_{(s)} \tag{17}$$

where, D is a constant matrix of dimension $(r \times 2NR)$; where, $NR = n + r$, and n is the number of degrees of freedoms (DOFs) of the structure, and r is the total number of SA-TMFDs. Furthermore, each damper displacement consists of two components.

$$y_{(s)} = y_{r(s)} + y_{b(s)} \tag{18}$$

where, $y_{r(s)}$ represents the slip displacement on the friction interfaces of the damper, while $y_{b(s)}$ represents the elastic deformation of the damper, which are proportional to the axial force of the damper. The axial force of the FD are equivalent to the friction force, therefore, by the elastic constitutive law for axial member, we have

$$F_{d(s)} = K_{Td} \cdot y_{b(s)} \tag{19}$$

where, K_{Td} is a $(r \times r)$ diagonal matrix consists of stiffness of the SA-TMFDs.

$$F_{d(s)} = K_{Td} [DZ_{(s)} - y_{r(s)}] \tag{20}$$

As it is clear from equation (20), the friction force vector $F_{d(s)}$ is a function of the current structural state $Z_{(s)}$ as well as the slip displacement on the friction interfaces of the two systems $y_{r(s)}$. At any given time instant each SA-TMFD unit of SA-MTMFDs can remain only in one state i.e. either in stick state or in slip state. During the time interval from $(s - 1)^{th}$ to $(s)^{th}$ time step, if each damper is in stick state then it should satisfy the following condition.

$$y_{r(s)} = y_{r(s-1)} \tag{21}$$

By applying the results of equations (20) and (21), the subtraction of $F_{d(s-1)}$ and $F_{d(s)}$ leads to

$$F_{d(s)} = K_{Td} D[Z_{(s)} - Z_{(s-1)}] + F_{d(s-1)} \tag{22}$$

Now, introducing equation (15) into equation (22) and replacing subscript (s) by $(s - 1)$ leads to

$$\tilde{F}_{d(s)} = G_z Z_{(s-1)} + G_{xg} \ddot{x}_{g(s-1)} + G_{fd} F_{d(s-1)} \tag{23}$$

where,

$$\left. \begin{aligned} G_z &= K_{Td} D(A_d - I) \\ G_{xg} &= K_{Td} DE_d \\ G_{fd} &= K_{Td} DB_d + I \end{aligned} \right\} \tag{24}$$

Note that in equation (23), $\tilde{F}_{d(s)}$ shows the damper forces computed by assuming that each damper is in stick state which may not be equal to actual friction force $F_{d(s)}$. As vector $\tilde{F}_{d(s)}$ plays a very important role in deciding the state (either stick or slip) and actual friction force in the damper. It shows the minimum friction force required by the damper at the t^{th} time step to remain in stick state and thus it is referred as ‘critical friction force’. Equation (23) shows that vector $\tilde{F}_{d(s)}$ can be computed easily, once $Z_{(s-1)}$, $F_{d(s-1)}$ and $\ddot{x}_{g(s-1)}$ have been determined at the previous time step. Further, it is assumed that damper obeys Coulomb’s friction law. In this case the actual friction force vector $F_{d(s)}$ and critical friction force vector $\tilde{F}_{d(s)}$ shall be reduce to scalars $F_{d(s)}$ and $\tilde{F}_{d(s)}$. The state of the damper can be decided to be

a) Stick state, if

$$|\tilde{F}_{d(s)}| < F_{d \max(s)} = F_c N_{(s)} \tag{25a}$$

b) Slip state, if

$$|\tilde{F}_{d(s)}| > F_{d \max(s)} = F_c N_{(s)} \tag{25b}$$

where, F_c is the friction coefficient and $N_{(s)}$ is the time varying clamping force of the damper. Using these equations, once the state of the damper is determined, its frictional force can be calculated by

$$\left. \begin{aligned} F_{d(s)} &= \tilde{f}_{d(s)} \quad (\text{for stick state}) \\ F_{d(s)} &= \text{sgn}[\tilde{f}_{d(s)}] F_c N_{(s)} \quad (\text{for slip state}) \end{aligned} \right\} \quad (26)$$

where, sgn denotes the signum function which takes the sign of variable and is used to denote the direction of the resisting slip force. Once $F_{d(s)}$ is obtained from equation (26) and substituted into equation (15), the structural response $Z_{(s+1)}$ can be determined and then the response of the system in next time step can be simulated.

Equation (25b) shows that if the clamping force $N_{(s)}$ is applied in such a way that resulting slip force is always slightly less than the value $\tilde{F}_{d(s)}$ predicted by equation (23) then the damper will remain in the slip state for the complete duration of the harmonic or earthquake excitation. Based on this concept, the control rule for determining the clamping force of a semi-active friction damper is proposed by Lu (2004) as

$$N_{(s)} = \alpha \frac{|\tilde{F}_{d(s)}|}{F_c}, \quad 0 \leq \alpha \leq 1 \quad (27)$$

where, α is a selectable constant parameter known as gain multiplier and $\tilde{F}_{d(s)}$ is obtained from equation (23), substituting $N_{(s)}$ from equation (27) into equation (25b), keeps the equation (25b) always true for each damper and keep each damper in its slip state. Therefore the damper friction forces can be computed by substituting equation (27) into equation (25) and re-writing it in a vector form as

$$F_{d(s)} = \alpha \tilde{F}_{d(s)} \quad (28)$$

Equation (28) shows that if the value of α is such as $0 \leq \alpha \leq 1$, damper friction force vector $F_{d(s)}$ will be always less than $\tilde{F}_{d(s)}$. By using equation (23) in equation (28), one can obtain an explicit formula to calculate the control forces vector as

$$F_{d(s)} = \alpha \left\{ G_z \cdot Z_{(s-1)} + G_{fd} \cdot F_{d(s-1)} + G_{xg} \cdot \ddot{x}_{g(s-1)} \right\} \quad (29)$$

From equation (29), it is noted that the parameter α plays an important role in the proposed predictive control law.

V. NUMERICAL STUDY

For the numerical study the five story shear type structure of fundamental time period of 0.5 sec is considered. The earthquake time histories along with their peak ground acceleration (PGA) and components which are used for this study are represented in Table 1. The displacement and acceleration response spectra of the above mentioned earthquakes are shown in Figure 2 for 2% critical damping. The maximum value of PGA are 1.225 g, 3.616 g, 3.296 g, 3.614 g, occurring at the period of 0.46 s, 0.64 s, 0.08 s and 0.36 s for Imperial Valley, Loma Prieta, Landers and Kobe earthquakes, respectively. The spectra of these ground motion indicate that these ground motions are recorded on a rocky site or on a firm soil. For the numerical study, the SA-MTMFDs are assumed to be attached to the top story of the structure as shown in Figure 1. The damping ratio of the primary system / structure is taken as 2%, constant for all modes of vibration. The weight of each floor is taken as 10000 kg. The natural frequencies of the structure are calculated as 2, 5.838, 9.203, 11.822, 13.484 Hz. The mass ratio, μ , is taken as 5% of the total weight of the primary system. For the present study, the results are obtained with the interval, $\Delta t = 0.02, 0.01$ and 0.005 , respectively. The number of iteration in each time step is taken as 50 to 200 to determine the incremental frictional force of the SA-TMFDs.

The controlling parameter α on which the efficiency of SA-TMFDs depends and the controlling parameter R_f on which the efficiency of P-MTMFDs depends, are discussed here for a fair comparison between the two. For the study purpose all the considered system parameters of the configured system attached with P-MTMFDs and SA-MTMFDs are kept same. The response quantities of the interest considered for the study are peak values of displacement, acceleration of the top story of the structure and average damper displacement of all the SA-MTMFDs.

5.1 Effect of Controlling Parameters on the Performance of P-MTMFDs and SA-MTMFDs

To investigate the performance of P-TMFDs and SA-MTMFDs in response reduction of five story shear structure, the optimum number of P-TMFD units and SA-TMFD units in P-TMFDs and SA-TMFDs are found out respectively. For this purpose the number of P-TMFD unit and SA-TMFD unit is varied as 1, 5 and 11. Also, for a fair comparison of performance of P-TMFDs and SA-TMFDs, optimum values of their respective controlling parameters are found out.

To determine the optimum value of controlling parameter R_f and study its influence on the performance of P-MTMFDs, the value of R_f (i.e. maximum friction force of the damper normalized by the weight of the P-TMFD) is varied from 1 to 50%. The variation of peak displacement, peak acceleration of the top story and the average stroke of P-MTMFDs is plotted against R_f in Figures 3, 4 and 5, respectively. It is observed from Figure 3 and 4 that as the value of R_f increases the peak displacement and peak acceleration response of the top story decreases up to a certain point and further increases gradually with the increase in the value of R_f . It is also, observed from Figure 5 that as the value of R_f increases the average stroke of P-MTMFDs decreases. Further, the study of peak response reduction with respect to variation in average peak stroke with respect to same value of R_f , shows that the peak responses of structure decreases with the decrease in the value of average peak stroke up to a certain point and again gradually increases with further reduction in the average peak stroke. Thus, giving emphasis on the maximum reduction of peak responses of the structure and reasonable value of average stroke, the optimum value of R_f is chosen. The optimum value of R_f at which the response of the system attains its minimum value is taken as 2%, 4%, 0.1% and 7%, for Imperial Valley, Loma Prieta, Landers and Kobe earthquakes, respectively. The variation of the optimum value of R_f for different earthquake is due to their different dynamic characteristic. It is also, observed from the figures that there exist an optimum number of TMFD units in P-MTMFDs at which P-TMFDs perform effectively and reduces the responses of system. Further, it is to be noted that increasing the number of TMFD units is not desirable from the economical point of view, once the optimum number of TMFD units are obtained. The optimum number of P-TMFD unit in P-MTMFDs are taken as 5 TMFD units for Imperial valley, Landers, Kobe earthquakes and 11 TMFD units for Loma Prieta earthquake.

Similarly, to find out the optimum value of controlling parameter α and study its influence on the performance of SA-MTMFDs, the value of α is varied from 0.1 to 0.999. The peak displacement, peak acceleration of the top story and the average stroke and average friction force (i.e. average of maximum friction forces of the dampers normalized by the weight of the SA-TMFDs) developed in the SA-TMFDs are plotted against α in Figure 6, 7, 8 and 9. It is observed from the Figures 6, 7, 8 and 9 that as the value of α increases the top story displacement, top story acceleration and the average peak stroke decreases and average frictional force of the SA-MTMFDs increases. Also, at a value of α which is extremely close to one, the peak displacement and peak acceleration of top story increases. Also, it is possible that at value of α which is extremely close to one, the SA-MTMFDs may enter in stick state for certain time instants, which may affect the energy dissipation capacity of the SA-MTMFDs. Hence by selecting an appropriate value of α , one can keep SA-TMFD continuously in slip mode and utilize its energy dissipation capacity effectively. Thus, for a given earthquake excitation an optimum value of α exist at which the response of the system attains its minimum value. The optimum value of α is chosen as 0.92, 0.9, 0.999 and 0.8, for Imperial Valley, Loma Prieta, Landers and Kobe earthquakes, respectively. It is also observed that an optimum value of number of SA-TMFD unit exists in SA-MTMFDs at which the SA-MTMFDs perform effectively. The optimum number of SA-TMFD unit in SA-MTMFDs is taken as 5 TMFD units for Imperial Valley, Landers, Kobe earthquakes and 11 TMFD units for Loma Prieta earthquake. Thus, from the above study, it is summarized as by selecting an appropriate value of α one can keep SA-MTMFDs continuously in slip mode and utilize its energy dissipation capacity effectively. For a given earthquake excitation an optimum value of α exists at which the response of the system attains its minimum value. The variation of the optimum value of α for different earthquake is due to their different dynamic characteristic. Further, an optimum value of number of SA-TMFD unit exists in SA-MTMFDs at which the SA-MTMFDs perform effectively.

5.2 Effects of PGA

In order to study the effect of PGA on the responses of interest, the PGA of earthquake time histories are scaled from 0.05 g to 1g. The peak displacement and peak acceleration of the top story along with the average of peak damper displacement i.e. average peak stroke of a MDOF system with P-MTMFDs, SA-MTMFDs and uncontrolled system are plotted against the different PGA level for various earthquakes in Figures 10, 11 and 12. For a fair comparison, the responses of P-MTMFDs are plotted with the optimum number of P-TMFD units and optimum value of R_f for each earthquake as obtained in earlier section. Also, the responses of SA-MTMFDs are plotted with optimum number of SA-TMFD units and optimum values of α . It is observed that both P-MTMFDs and SA-MTMFDs reduce the response of interest effectively. Even for some earthquakes optimally selected P-MTMFDs and SA-TMFDs performs at par. However for most of the earthquakes the response reduction ability of SA-TMFDs is higher than that of P-MTMFDs. It is also observed from the figure that sometimes at very low PGA levels like 0.05 g and 0.1 g the value of average peak stroke of P-MTMFDs is close to zero, which shows that at very low intensity earthquakes the P-MTMFDs hardly activates or underperforms. It is also observed that the SA-MTMFDs can be activated at all PGA levels and is also effective in reducing the response of the MDOF system at all PGA levels, due to this SA-MTMFDs overcomes all the limitations of P-MTMFDs.

In a similar manner, Figures 13, 14, 15 and 16 depict the displacement and acceleration time history of top story of the primary system attached with P-MTMFDs, SA-MTMFDs and uncontrolled system for optimum value of R_f and α with optimum number of their respective TMFD units. For this purpose, the PGA of all the considered earthquakes is scaled to 0.4 g and 0.9 g, which shows the low and high intensity level earthquakes, respectively. The time history responses of the system confirms that the SA-MTMFDs are more effective than P-MTMFD in response reduction of the MDOF system as it is activated at such a lower and higher PGA levels.

5.3 Effect of Variation of Mass Ratio, Tuning Ratio and Frequency Spacing

Figures 17, 18 and 19 depict the effectiveness of control algorithm, when assuming the changes in the parameters or properties of the P-MTMFDs and SA-MTMFDs. For this purpose, the responses of P-MTMFDs and SA-MTMFDs is plotted against the varying mass ratio, tuning ratio and frequency spacing in Figures 17, 18 and 19, respectively. It is observed that the response of the system is relatively less sensitive to the change in mass ratio of the system. While, in case of change in the tuning property and frequency spacing of the system, it is more sensitive. It is also observed that the responses of interest are more sensitive for the SA-MTMFDs in compare to responses of P-MTMFDs having values of responses at lower side in compare to P-MTMFDs. So, even if the actual friction force applied at SA-MTMFDs is different (due to change in properties/ parameters like mass ratio, tuning ratio and frequency spacing of SA-MTMFDs) than that of the friction force calculated from Predictive control law, SA-MTMFDs slightly alters the responses of the system at lower side in compare to its passive counterpart, which ensures its better performance level in compare to P-MTMFDs.

VI. CONCLUSIONS

The response of five story shear type structure attached with P-MTMFDs and SA-MTMFDs is investigated under four different seismic excitations. The predictive control law proposed by Lu (2004) is used for this study as it produces continuous and smooth slip force throughout the duration of an excitation. The governing differential equations of motion of the system are solved numerically, using state space method, to find out the response of the system. To investigate the effectiveness of SA-MTMFDs with predictive control, the responses of the system with P-MTMFDs are compared with the responses of the system with SA-MTMFDs. On the basis of trends of results obtained, the following conclusions are drawn:

1. By selecting an appropriate value of α one can keep SA-MTMFDs continuously in slip mode and utilize its energy dissipation capacity effectively.
2. For a given earthquake excitation an optimum value of α exists at which the response of the system attains minimum value. The variation of the optimum value of α for different earthquake is due to their different dynamic characteristic.
3. An optimum value of number of SA-TMFD unit exists in SA-MTMFDs at which the SA-MTMFDs perform effectively.
4. SA-MTMFDs can be activated at all PGA levels and is also effective in reducing the response of the MDOF system at all PGA levels, due to this SA-MTMFDs overcomes all the limitations of P-MTMFDs.

- If the actual friction force applied at SA-MTMFDs is different (due to change in parameters like mass ratio, tuning ratio and frequency spacing of SA-MTMFDs) than that of the friction force calculated from predictive control law, SA-MTMFDs slightly alters the responses of the system at lower side in compare to its passive counterpart, which ensures its better performance level in compare to P-MTMFDs.

Earthquake	Recording Station	Component	Duration (Sec)	PGA (g)
Imperial Valley (19 th May 1940)	El Centro Array # 9	I – ELC 180	40	0.313
Loma Prieta (18 th October 1989)	UCSC 16 LOS Gatos Presentation Centre (LGPC)	LGP 000	25	0.96
Landers 28 th June 1992	Lucerene Valley	LCN 275	48.125	0.721
Kobe 16 th January 1995	Japan Meterological Agency (JMA) 99999 KJMA	KJM 000	48	0.82

Table 1: Details of Earthquakes considered for Numerical study.

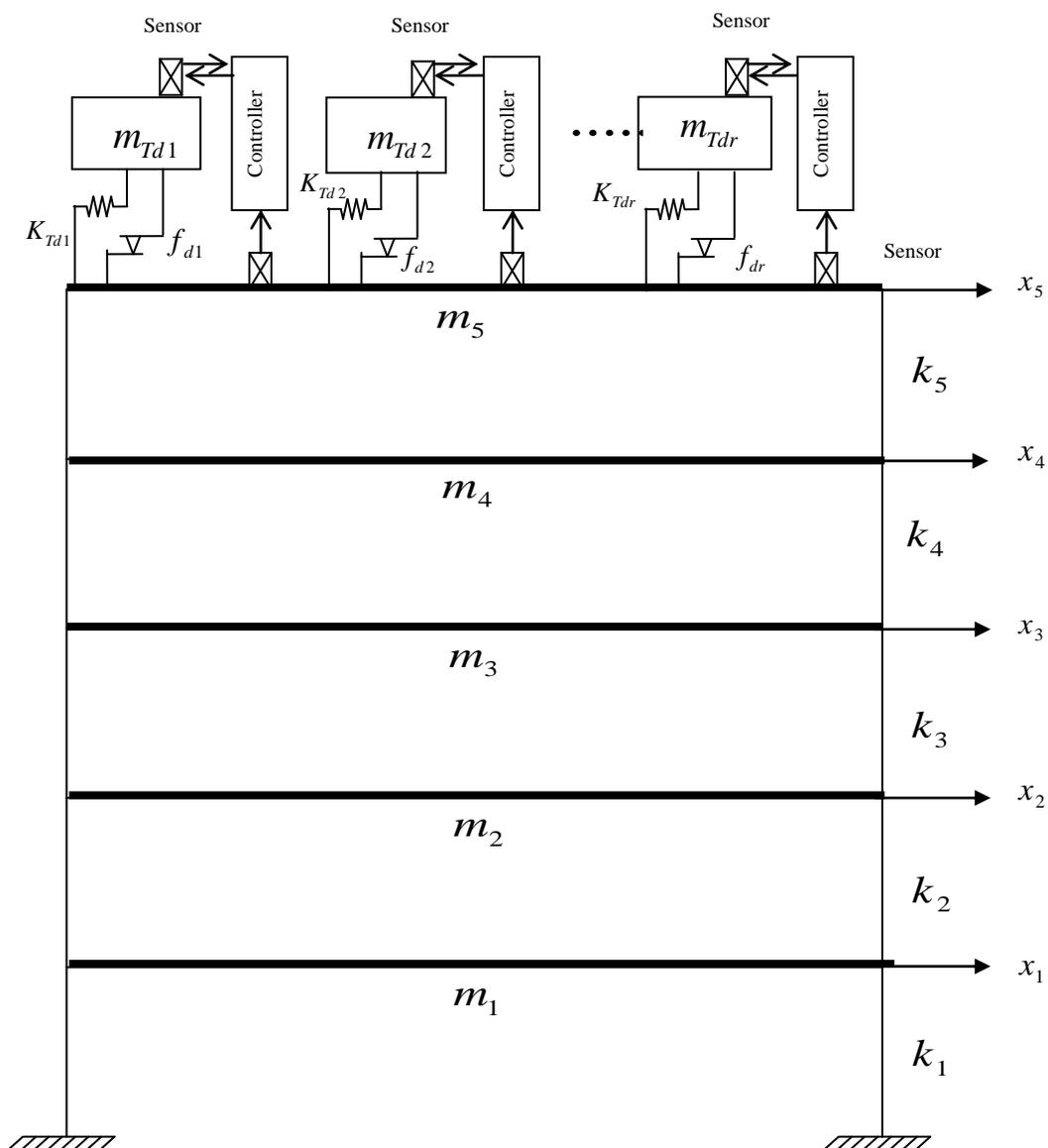


Figure 1: Schematic diagram of a multi-story structure with semi-active multiple tuned mass friction dampers.

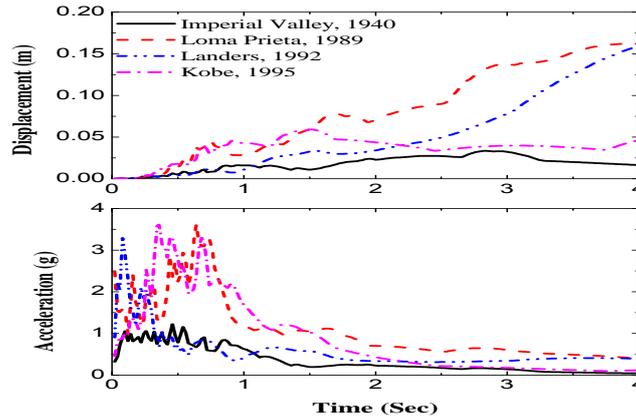


Figure 2: Displacement and acceleration spectra of the selected earthquakes.

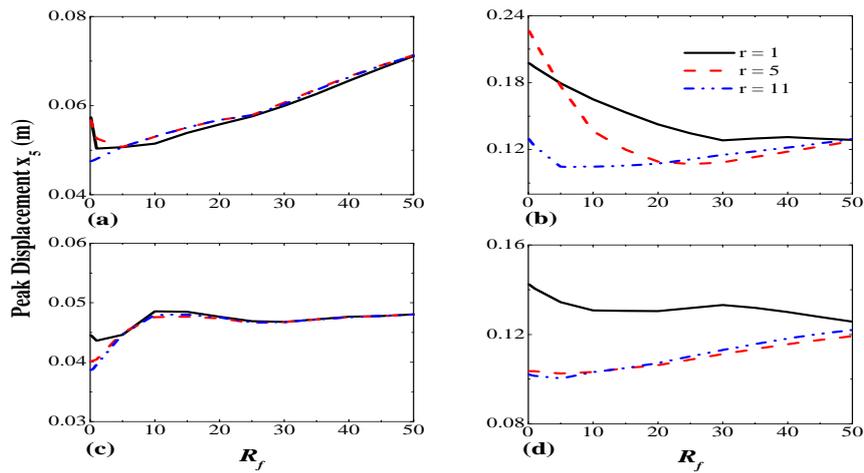


Figure 3: Influence of R_f on peak displacement on response of P-MTMFDs. (a) Imperial Valley, 1940; (b) Loma Prieta, 1989; (c) Landers, 1992; (d) Kobe, 1995.

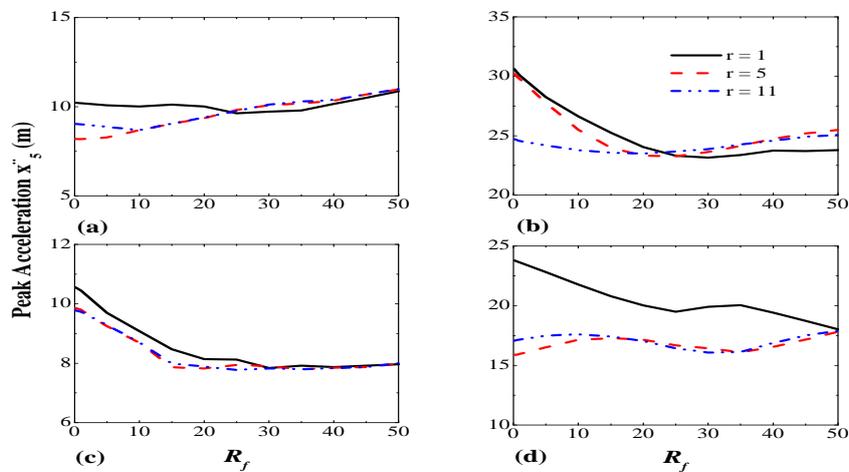


Figure 4: Influence of R_f on peak acceleration response of P-MTMFDs. (a) Imperial Valley, 1940; (b) Loma Prieta, 1989; (c) Landers, 1992; (d) Kobe, 1995.

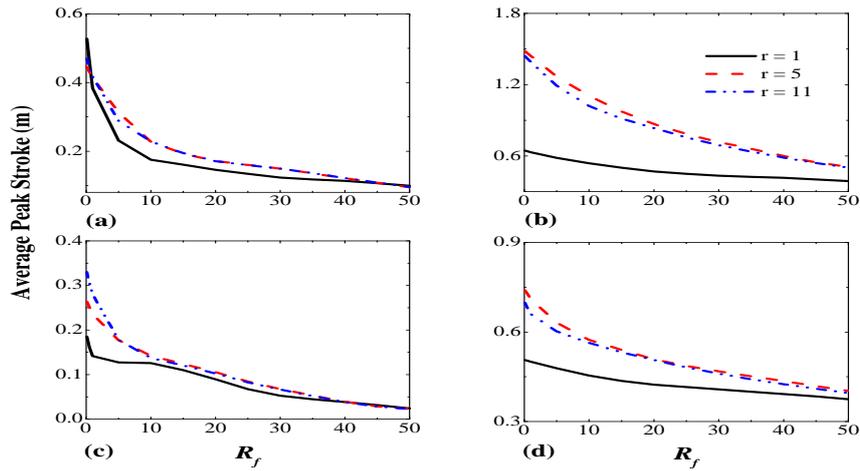


Figure 5: Influence of R_f on average peak stroke response of P-MTMFDs. (a) Imperial Valley, 1940; (b) Loma Prieta, 1989; (c) Landers, 1992; (d) Kobe, 1995.

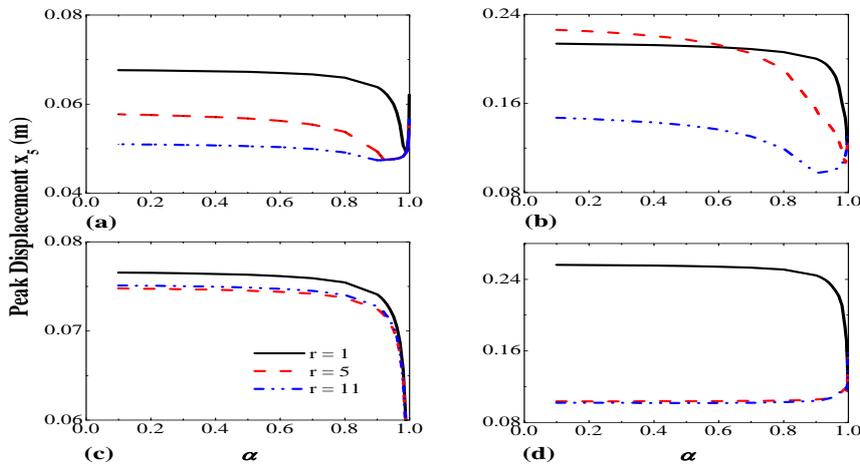


Figure 6: Influence of α on peak displacement response of SA-MTMFDs. (a) Imperial Valley, 1940; (b) Loma Prieta, 1989; (c) Landers, 1992; (d) Kobe, 1995.

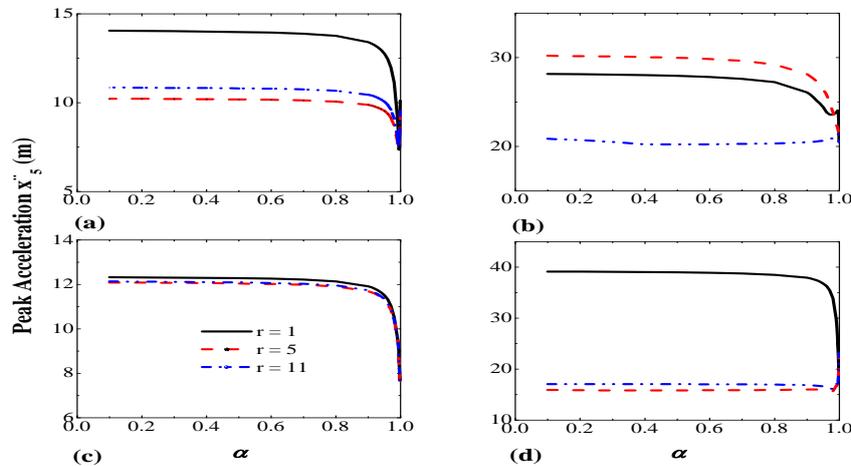


Figure 7: Influence of α on peak acceleration response of SA-MTMFDs. (a) Imperial Valley, 1940; (b) Loma Prieta, 1989; (c) Landers, 1992; (d) Kobe, 1995.

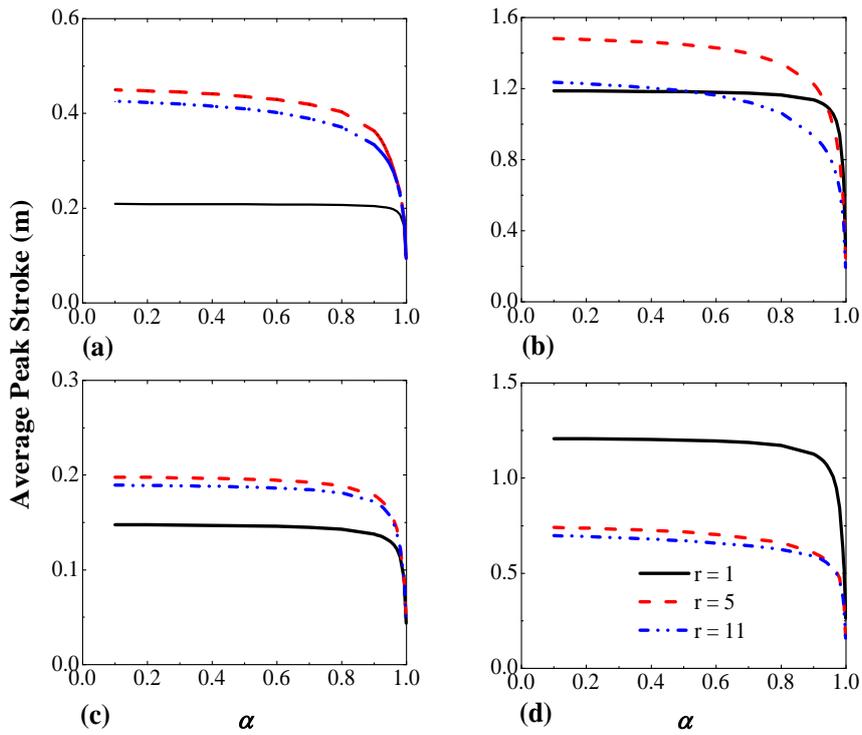


Figure 8: Influence of α on average peak stroke response of SA-MTMFDs. (a) Imperial Valley, 1940; (b) Loma Prieta, 1989; (c) Landers, 1992; (d) Kobe, 1995.

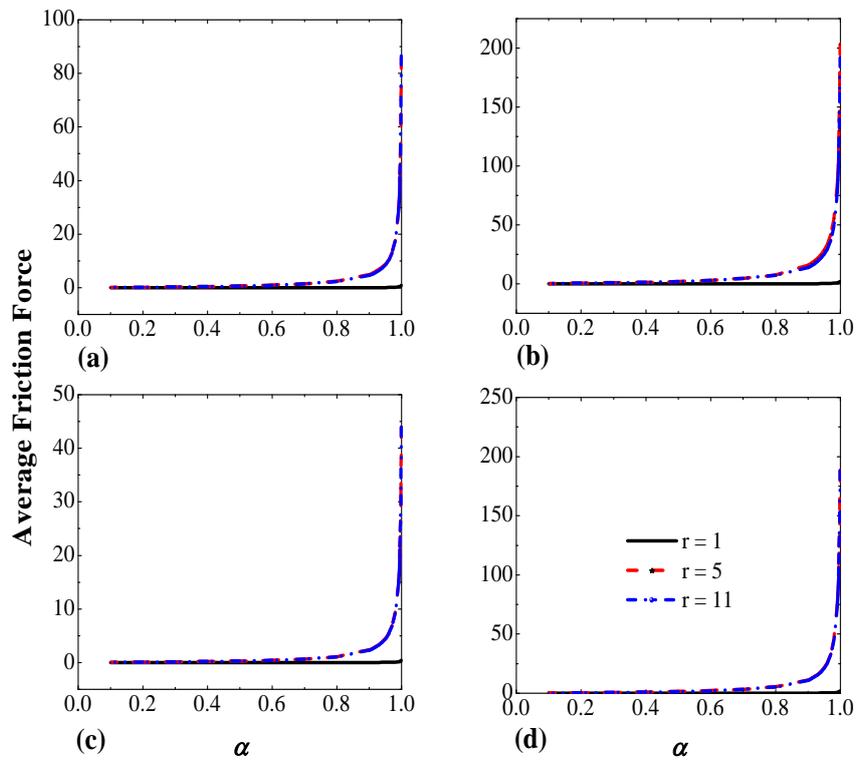


Figure 9: Influence of α on average friction force of SA-MTMFDs. (a) Imperial Valley, 1940; (b) Loma Prieta, 1989; (c) Landers, 1992; (d) Kobe, 1995.

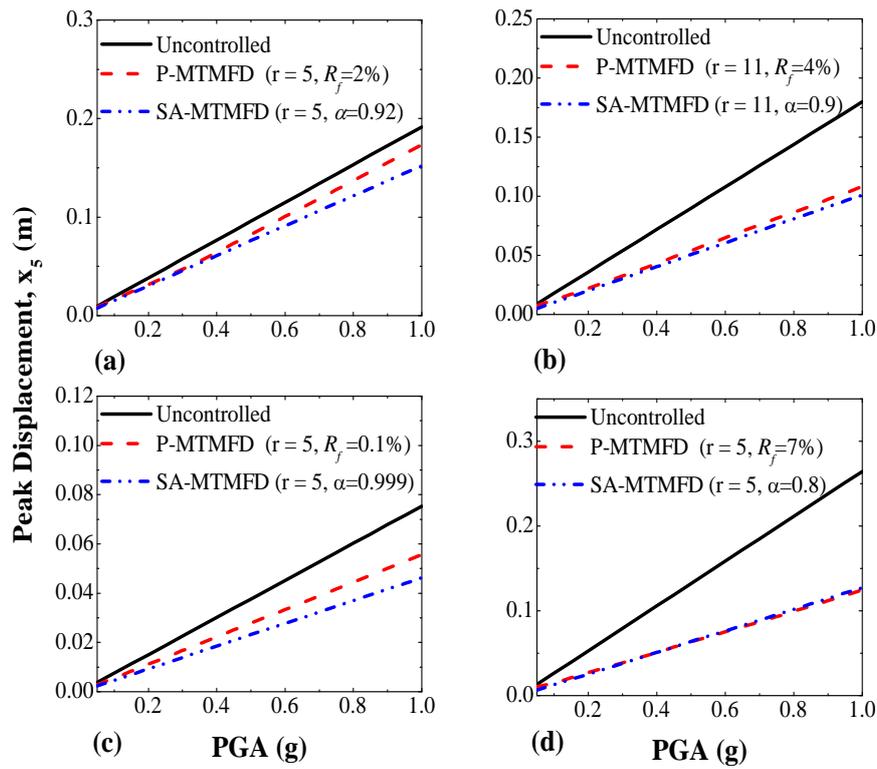


Figure 10: Peak displacement responses of P-MTMFDs and SA-MTMFDs under different Earthquakes. (a) Imperial Valley, 1940; (b) Loma Prieta, 1989; (c) Landers, 1992; (d) Kobe, 1995.

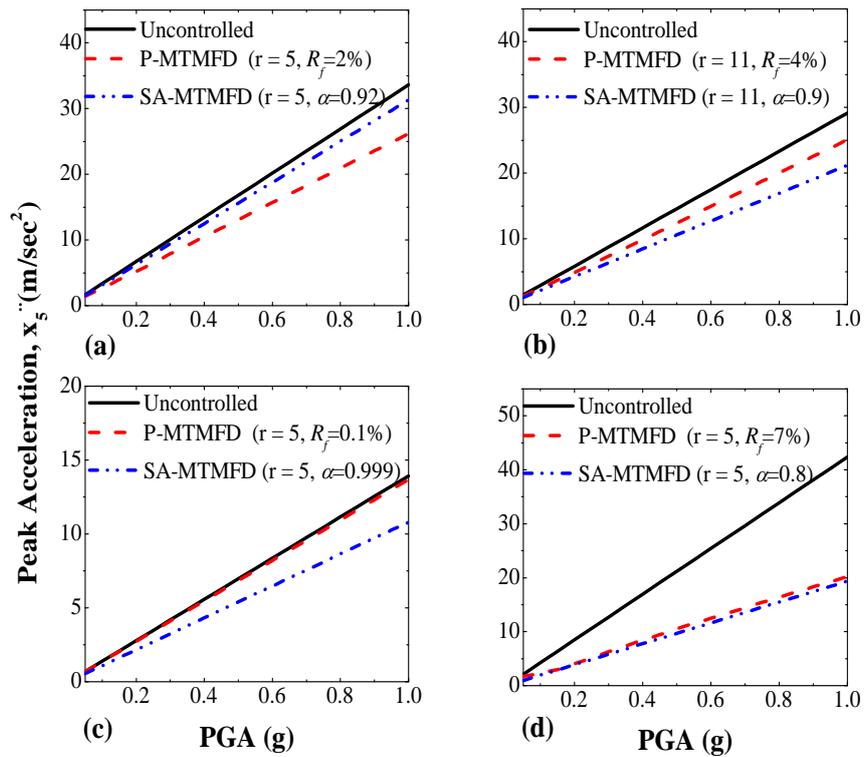


Figure 11: Peak acceleration responses of P-MTMFDs and SA-MTMFDs under different Earthquakes. (a) Imperial Valley, 1940; (b) Loma Prieta, 1989; (c) Landers, 1992; (d) Kobe, 1995.

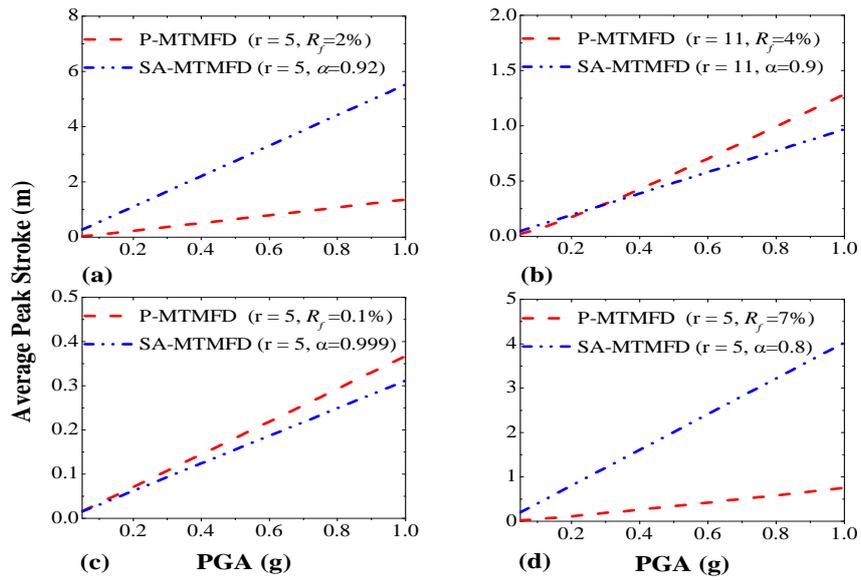


Figure 12: Peak average stroke responses of P-MTMFDs and SA-MTMFDs under different Earthquakes. (a) Imperial Valley, 1940; (b) Loma Prieta, 1989; (c) Landers, 1992; (d) Kobe, 1995.

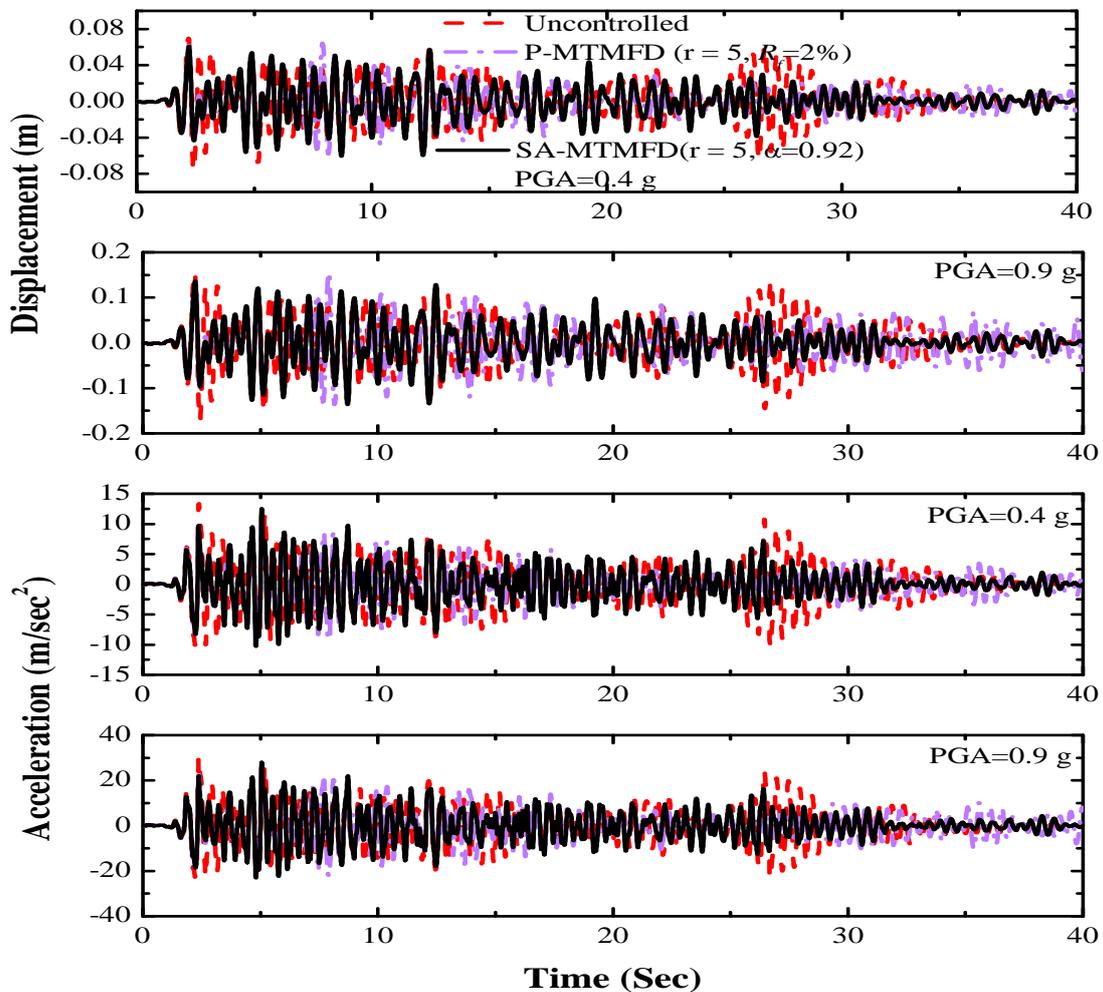


Figure 13: Comparison of Displacement & Acceleration responses of Uncontrolled, P-MTMFDs and SA-MTMFDs for Imperial Valley Earthquake (1940) for PGA as 0.4 g and 0.9 g.

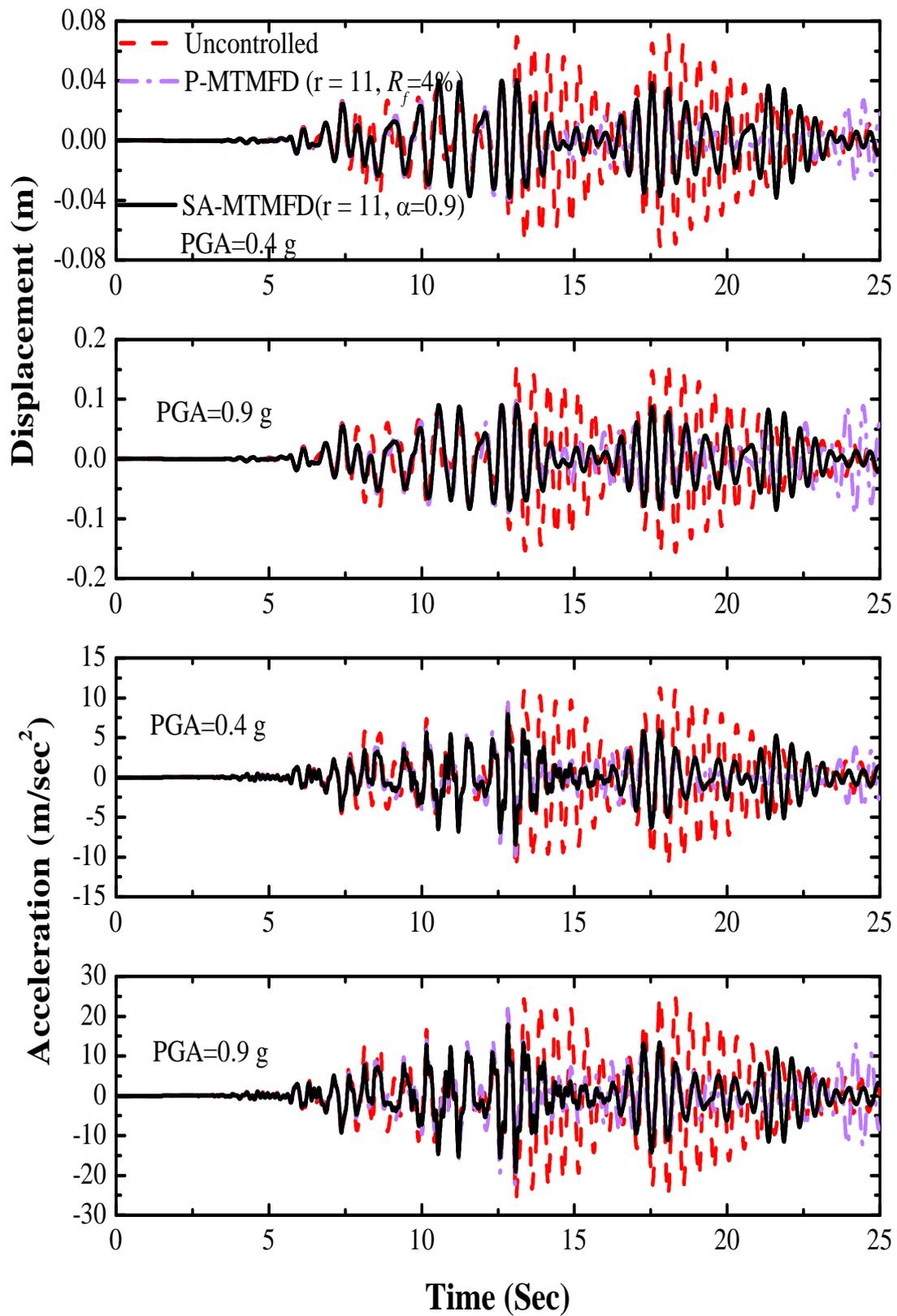


Figure 14: Comparison of Acceleration & Displacement responses of Uncontrolled, P-MTMFDs and SA-MTMFDs for Loma Prieta Earthquake (1989) for PGA as 0.4 g and 0.9 g.

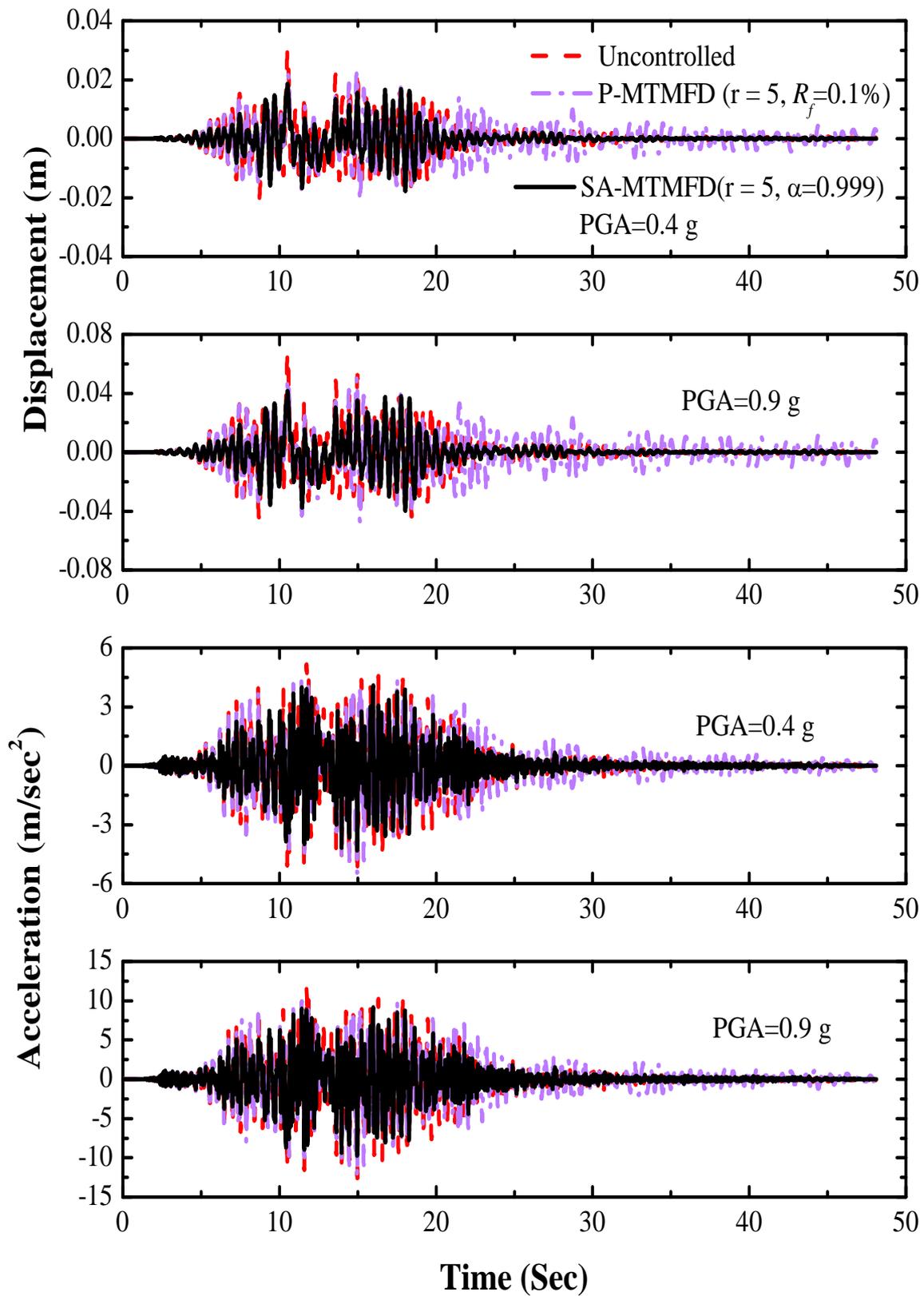


Figure 15: Comparison of Acceleration & Displacement responses of Uncontrolled, P-MTMFDs and SA-MTMFDs for Landers Earthquake (1992) for PGA as 0.4 g and 0.9 g.

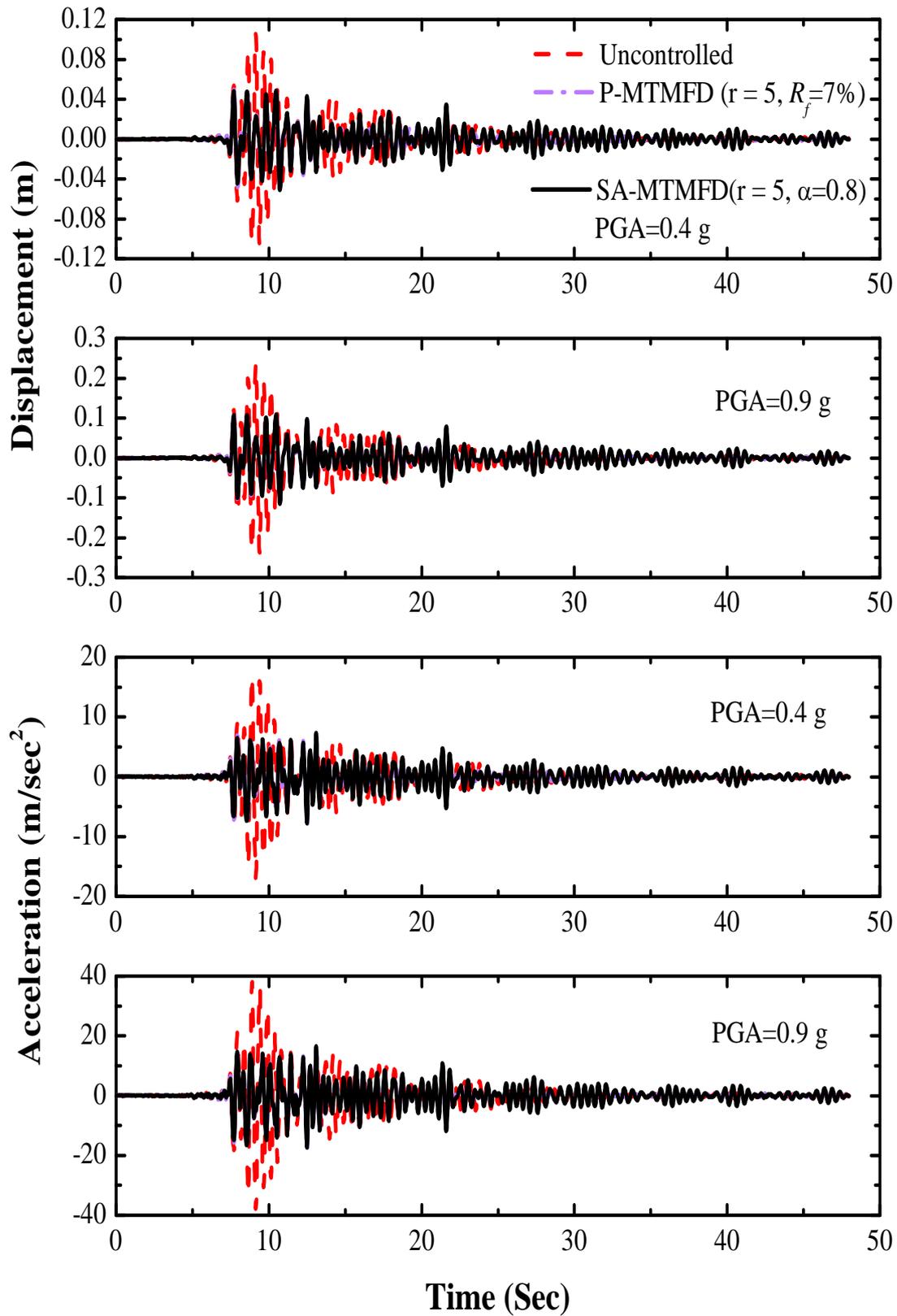


Figure 16: Comparison of Acceleration & Displacement responses of Uncontrolled, P-MTMFDs and SA-MTMFDs for Kobe Earthquake (1994) for PGA as 0.4 g and 0.9 g.

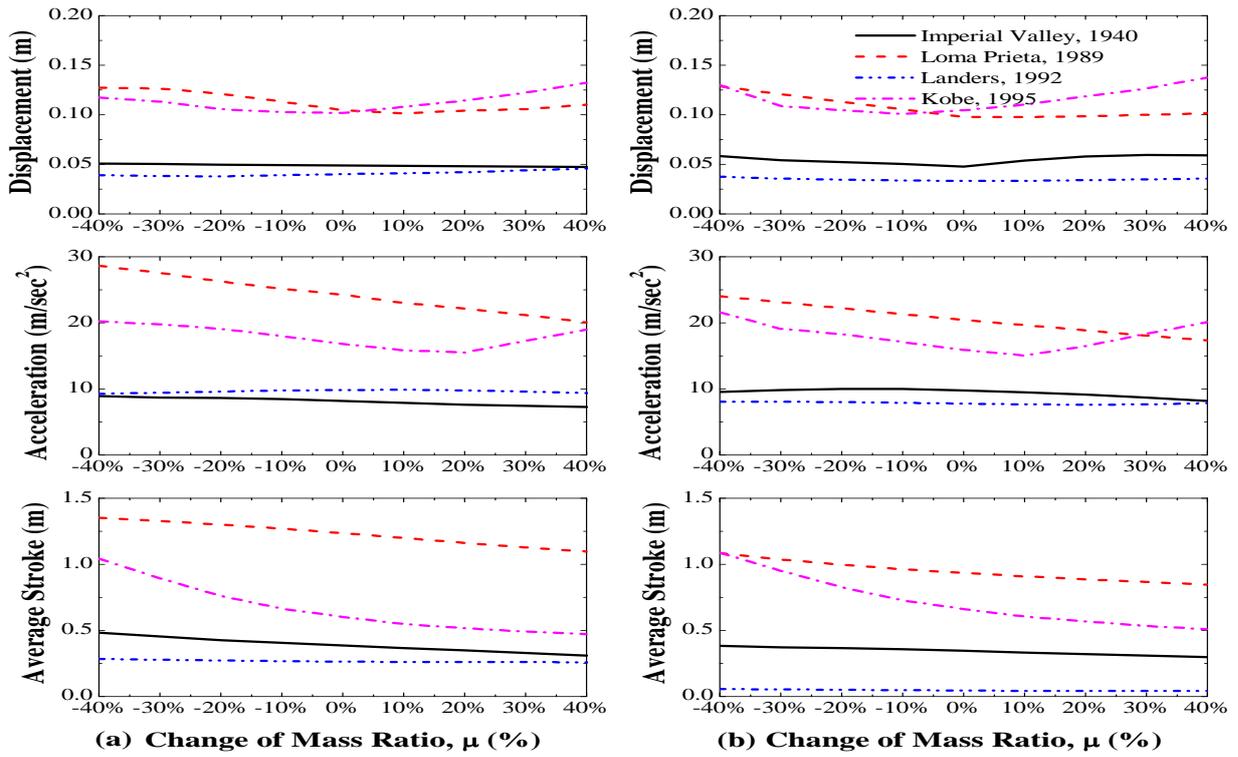


Figure 17: Effect of percentage variation in the mass ratio of P-MTMFDs and (b) SA-MTMFDs.

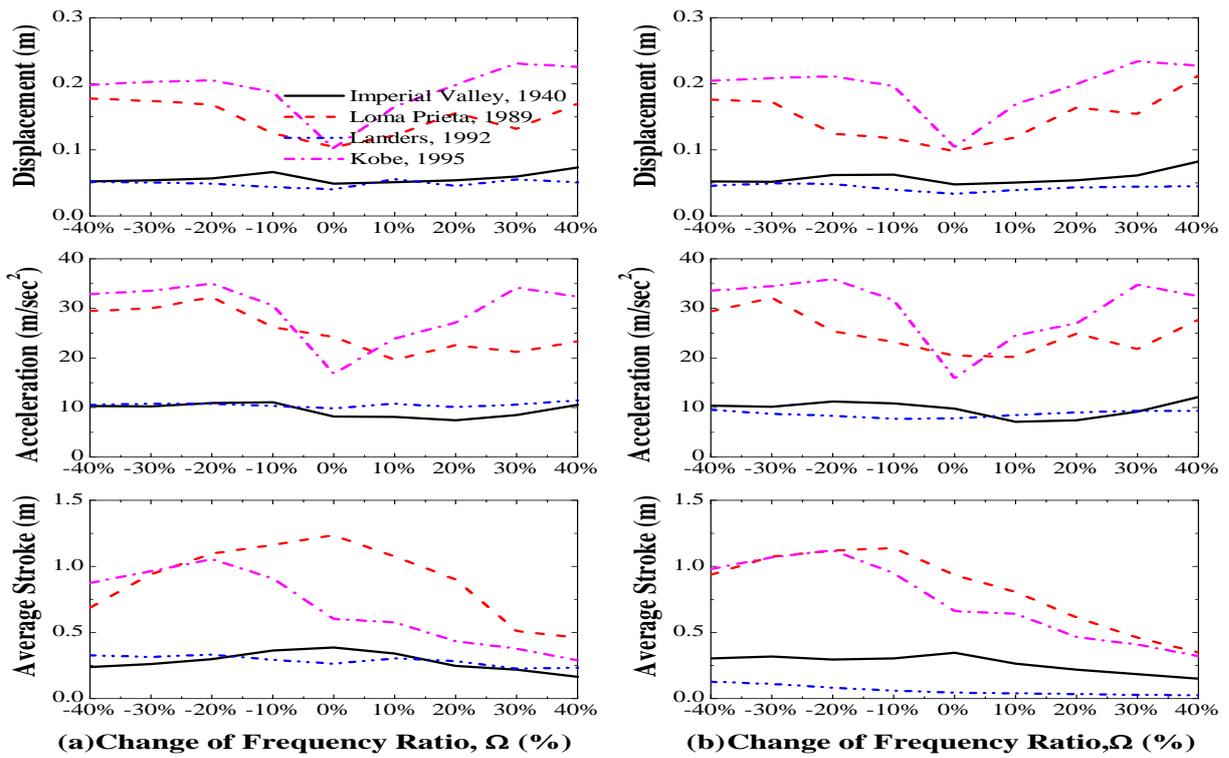


Figure 18: Effect of percentage variation in the frequency ratio of (a) P-MTMFDs and (b) SA-MTMFDs.

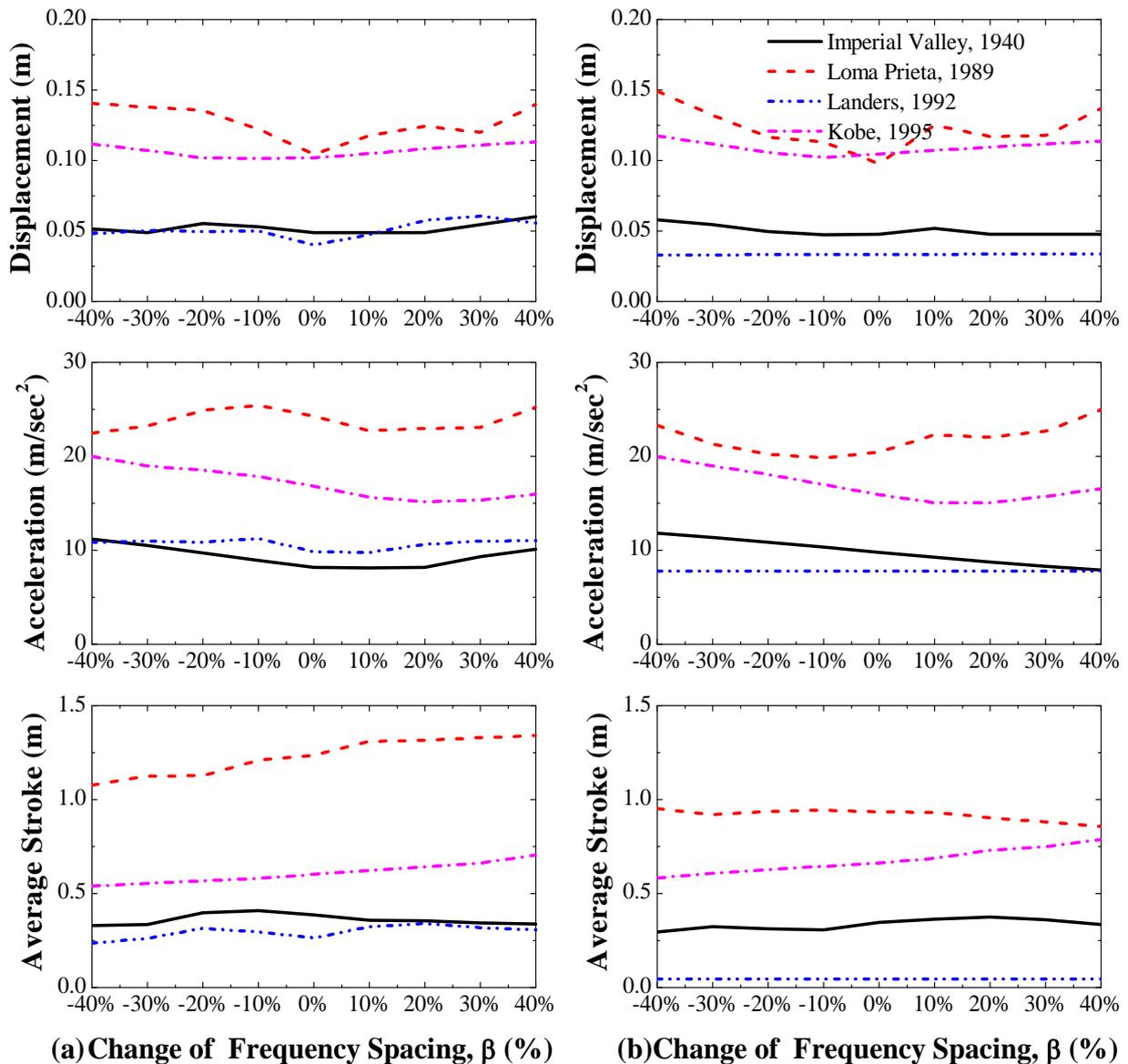


Figure 19: Effect of percentage variation in the frequency spacing of (a) P-MTMFDs and (b) SA-MTMFDs.

REFERENCES

- [1] Z. Akbay and H.M. Aktan. Abating earthquake effects on buildings by active slip brace devices. *Shock and Vibration*, 2(2), 133-142, 1995.
- [2] P. Colajanni and M. Papia. Seismic response of braced frames with and without friction dampers. *Engineering Structures*, 17, 129-40, 1995.
- [3] D.J. Dowdell and S. Cherry. Structural control using semi-active friction dampers. *Proceedings of the First World Conference on Structural Control*, Los Angeles, Vol. 3; FAI-59-68, 1994.
- [4] Z. Gewei and B. Basu. A Study on friction-tuned mass damper: Harmonic solution and statistical linearization. *Journal of Vibration and Control*, 17, 721-731, 2010.
- [5] J.A. Inaudi. Modulated homogeneous friction: A semi-active damping strategy. *Earthquake Engineering and Structural Dynamics*, 26(3), 361-376, 1997.
- [6] A.S. Joshi and R.S. Jangid. Optimum parameters of multiple tuned mass dampers for base excited damped systems. *Journal of Sound and Vibration*, 202, 657-667, 1997.
- [7] S. Kannan, H.M. Uras and H.M. Aktan. Active control of building seismic response by energy dissipation. *Earthquake Engineering and Structural Dynamics*, 24(5), 747-759, 1995.
- [8] J.H. Lee, E. Berger and J.H. Kim. Feasibility study of a tunable friction damper *Journal of Sound and Vibration*, 283, 707-722, 2005.

- [9] C.C. Lin, G.L. Ling and J.F. Wang. Protection of seismic structures using semi-active friction TMD. *Earthquake Engineering and Structural Dynamics*, 39, 635-659, 2010.
- [10] L.Y. Lu. Predictive control of seismic structures with semi-active friction dampers. *Earthquake Engineering and Structural Dynamics*, 33(5), 647-668, 2004.
- [11] L. Meirovitch. *Dynamics and control of structures*, John Wiley & Sons, Inc., New York, 1990.
- [12] I.H. Mualla and B. Belev. Performance of steel frames with a new friction damper device under earthquake excitation. *Engineering Structures*, 24, 365-71, 2002.
- [13] C. Pasquin, N. Leboeuf, R.T. Pall and A.S. Pall. Friction dampers for seismic rehabilitation of Eaton's building. Montreal, 13th World Conference on Earthquake Engineering, Paper No. 1949, 2004.
- [14] W.L. Qu, Z.H. Chen and Y.L. Xu. Dynamic analysis of wind-excited truss tower with friction dampers. *Computers and Structures*, 79, 2817-31, 2001.
- [15] F. Ricciardelli and B.J. Vickery. Tuned vibration absorbers with dry friction damping. *Earthquake Engineering and Structural Dynamics*, 28, 707-723, 1999.
- [16] K. Xu and T. Igusa. Dynamic characteristics of multiple substructures with closely spaced frequencies. *Earthquake Engineering and Structural Dynamics*, 21(12), 1059-1070, 1992.

Overview Applications of Data Mining In Health Care: The Case Study of Arusha Region

¹Salim Diwani, ²Suzan Mishol, ³Daniel S.Kayange, ⁴Dina Machuve And ⁵Anael Sam

1,2,3,4,5,School of Computational and Communications Science and Engineering, Nelson Mandela-African Institution of Science and Technology (NM-AIST), Arusha, Tanzania.

ABSTRACT

Data mining as one of many constituents of health care has been used intensively and extensively in many organizations around the globe as an efficient technique of finding correlations or patterns among dozens of fields in large relational databases to results into more useful health information. In healthcare, data mining is becoming increasingly popular and essential. Data mining applications can greatly benefits all parties involved in health care industry. The huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods. Data mining provides the methodology and technology to transform huge amount of data into useful information for decision making. This paper explores data mining applications in healthcare in Arusha region of Tanzania more particularly; it discusses data mining and its applications in major areas such as evaluation of treatment effectiveness, management of healthcare itself and lowering medical costs.

KEYWORDS: Data mining, Healthcare application, Knowledge discovery, Data warehousing.

I. INTRODUCTION

Data mining can be defined as the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. Alternatively, it can be defined as the process of data selection and exploration and building models using massive data stores to uncover previously unknown patterns. Data mining is an analytic process designed to explore large amounts of data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. Data mining is not new idea, it has been used intensively and extensively by financial institutions for activities such as credit scoring and fraud detection[1, 2]; marketers for direct marketing and cross-selling; retailers for market segmentation and store layout; manufacturers for quality control and maintenance scheduling and it has been used in hospital care as well[2].Data mining has been becoming increasingly popular, it has been noted that several factors have motivated the use of data mining applications. The existence of medical insurance fraud and abuse, for example has led many healthcare insurers to attempt to reduce their losses by using data mining tools, the application has helped to help them find and track offenders. However fraud detection using data mining applications is prevailing in the commercial world for detection of fraudulent credit card transactions and fraudulent banking activities [3]. Huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods hence calls for technological interventions so as to simplify management of those data. Data mining can improve decision making by discovering patterns and trends in large amounts of complex data. Such analysis has become increasingly essential as financial pressures have amplified the need for healthcare organizations to make decisions based on the analysis of clinical and financial data. Insights gained from data mining can influence cost, revenue and operating efficiency while maintaining a high level of care.Healthcare organizations that perform data mining are better positioned to meet their long term needs; data can be a great asset to healthcare organizations, but they have to be first transformed into informationYet another factor motivating the use of data mining applications in healthcare is the realization that data mining can generate information that is very useful to all parties involved in the healthcare industry. For example, data mining applications can help healthcare insurers detect fraud and abuse, and healthcare providers can gain assistance in making decisions.. Data mining applications also can benefit healthcare providers such as hospitals, clinics, physicians, and patients by identifying effective treatments and best practices [4].

Data mining can be defined as the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. Alternatively, it can be defined as the process of data selection and exploration and building models using massive data stores to uncover previously unknown patterns. Data mining is an analytic process designed to explore large amounts of data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. Data mining is an automated approach for discovering or inferring hidden patterns or knowledge buried in data. 'Hidden' means patterns that are not made apparent through casual observation[5]. Data Mining is an interdisciplinary field that combines artificial intelligence, computer science, machine learning, database management, data visualization, mathematical algorithms, and statistics. Data Mining is a technology for knowledge discovery in databases (KDD). This technology provides different methodologies for decision making, problem solving, analysis, planning, diagnosis, detection, integration, prevention, learning and innovation[6]. Data mining is a variety of techniques such as neural networks, decision trees or standard statistical techniques to identify nuggets of information or decision making knowledge in bodies of data, and extracting these in such a way that they can be put to use in areas such as decision support, prediction, forecasting, and estimation.

Data mining has been becoming increasingly popular, it has been noted that several factors have motivated the use of data mining applications. The existence of medical insurance fraud and abuse, for example has led many healthcare insurers to attempt to reduce their losses by using data mining tools, the application has helped to help them find and track offenders. However fraud detection using data mining applications is prevailing in the commercial world for detection of fraudulent credit card transactions and fraudulent banking activities [7]. Huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods hence calls for technological interventions so as to simplify management of those data. Data mining can improve decision-making by discovering patterns and trends in large amounts of complex data. Such analysis has become increasingly essential as financial pressures have amplified the need for healthcare organizations to make decisions based on the analysis of clinical and financial data. Insights gained from data mining can influence cost, revenue and operating efficiency while maintaining a high level of care. Yet another factor motivating the use of data mining applications in healthcare is the realization that data mining can generate information that is very useful to all parties involved in the healthcare industry. For example, data mining applications can help healthcare insurers detect fraud and abuse, and healthcare providers can gain assistance in making decision.

Data mining applications can benefit healthcare providers in Arusha regions such as hospitals, clinics, physicians, and patients by adopting new technologies, which will help in early detection of life threatening diseases and lowering the medical costs. The aims of quality healthcare services are:

- providing safe healthcare treatments
- using scientific medical knowledge to provide healthcare services to everyone
- providing various healthcare treatments based on the patient's needs, symptoms and preferences
- minimizing the time to wait for the medical treatment
- minimizing the delay time in providing medical treatment.
- Health determinants
- Inputs to the health system and related processes (e.g., health infrastructure, human and financial resources, equipment, policy, and organization)
- Health outcomes (e.g., mortality, morbidity, disability, well-being, and health status)

II. DATA MINING

Nowadays there is huge amount of data stored in real-world databases and this amount continues to grow fast as it creates both an opportunity and a need for semi-automatic methods that discover the hidden knowledge in such database. If such knowledge discovery activity is successful, discovered knowledge can be used to improve the decision making process of an organization. For instance data about a hospital's patient might contain interesting knowledge about which kind of patient is more likely to develop a given disease. This knowledge can lead to better diagnosis and treatment for future patients[8]. Data mining and knowledge discovery is the name often used to refer to a very interdisciplinary field, which consists of using methods of several research areas to extract knowledge from real world data sets. There is a distinction between the terms data mining and knowledge discovery; the term data mining refers to the core steps of a broader process called knowledge discovery in database. In addition to the data mining step which actually extracts knowledge from data, the knowledge discovery process includes several preprocessing and post processing steps. The goal of data preparation methods is to transform the data to facilitate the application of a given data mining algorithms,

where the goal of knowledge refinement methods is to validate and refine discovered knowledge. The knowledge discovery is both iterative and interactive. It is iterative because the output of each step is often feedback to previous steps and typically many iterations of this process are necessary to extract high quality knowledge from data. It is interactive because the user or more precisely an expert in the application domain should be involved in this loop to help in data preparation, discovered-knowledge validation and refinement.

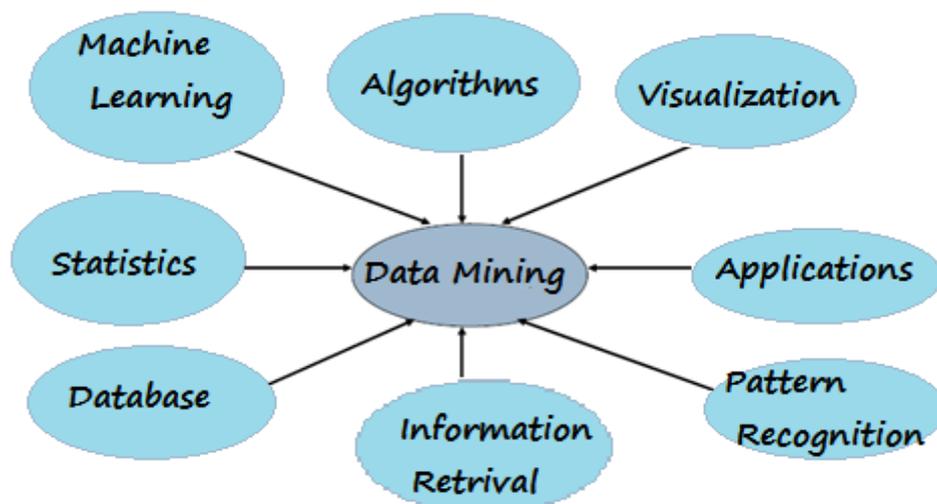


Fig 1: Data Mining Architecture

III. HEALTHCARE DATA MINING APPLICATIONS

There is vast potential for data mining applications in healthcare particularly in Arusha health centers. This is due to the fact that the use of technology can stand to provide accurate and more meaningful statistics of different activities going on within health centers. Generally, these activities can be grouped as the evaluation of treatment effectiveness, management of healthcare itself and customer relationship management[9]. Treatment effectiveness: Data mining applications can be developed to evaluate the effectiveness of medical treatments. By comparing and contrasting causes, symptoms, and courses of treatments, data mining can deliver an analysis of which courses of action prove effective,[10] for example the outcomes of patient groups treated with different drug regimens for the same disease or condition can be compared to determine which treatments work best and are most cost-effective. Other data mining applications related to treatments include associating the various side-effects of treatment, collating common symptoms to aid diagnosis, determining the most effective drug compounds for treating sub-populations that respond differently from the mainstream population to certain drugs and determining proactive steps that can reduce the risk of affliction. Future needs of individuals to improve their level of satisfaction[11]. These applications also can be used to predict other products that a healthcare customer is likely to purchase, whether a patient is likely to comply with prescribed treatment or whether preventive care is likely to produce a significant reduction in future utilization.

IV. KNOWLEDGE DISCOVERY

Knowledge discovery is a non-tedious procedure for identifying effective and potential benefits amid data. It is known from Fig. 2 below that data mining is one of the important processes of knowledge discovery. From the definitions by the scholars, it is clear that the usage of data mining is an analysis process within a series of knowledge discovery[12]. As time changes the term “data mining” gradually replaces “knowledge discovery”. From above, purpose of data mining is to uncover the rules that are helpful to decision process from massive data.

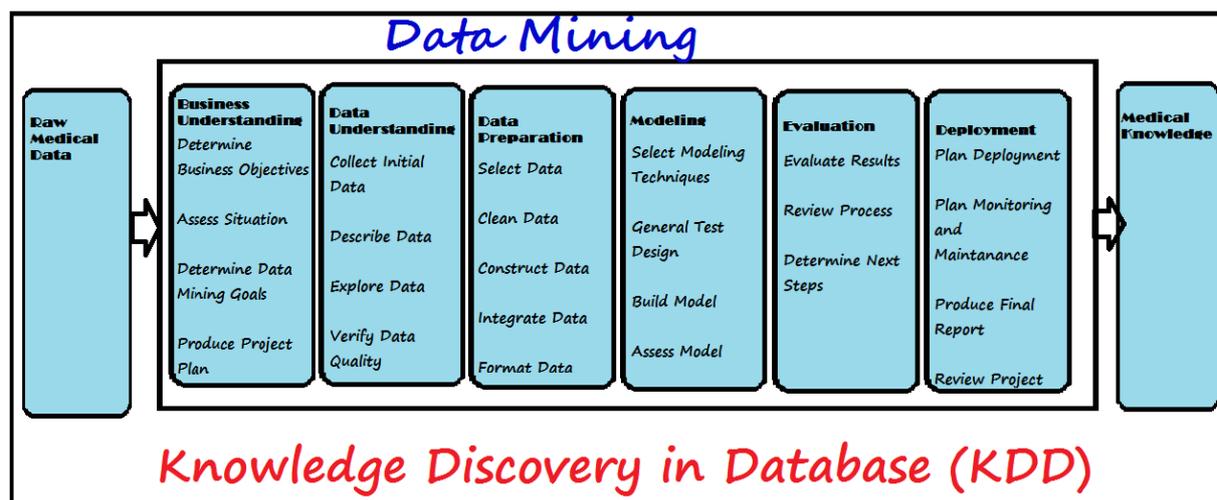


Fig 2: Knowledge Discovery

Knowledge Discovery As A Process

1. Data cleaning (to remove noise and inconsistent data)
 2. Data integration (where multiple data sources may be combined)
 3. Data selection (where data relevant to the analysis task are retrieved from the database)
 4. Data transformation (where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance)
 5. Data mining (an essential process where intelligent methods are applied in order to extract data patterns)
 6. Pattern evaluation (to identify the truly interesting patterns representing knowledge based on some interestingness measures)
 7. Knowledge presentation (where visualization and knowledge representation techniques are used to present the mined knowledge to the user)
- Steps 1 to 4 are different forms of data preprocessing, where the data are prepared for mining. The data mining step may interact with the user or a knowledge base. The interesting patterns are presented to the user and may be stored as new knowledge in the knowledge base. Note that according to this view, data mining is only one step in the entire process, albeit an essential one because it uncovers hidden patterns for evaluation.

We agree that data mining is a step in the knowledge discovery process. However, in industry, in media, and in the database research milieu, the term data mining is becoming more popular than the longer term of knowledge discovery from data.

Data cleaning and data integration techniques may be performed on the data. The database or data warehouse server is responsible for fetching the relevant data, based on the user's data mining request.

V. LIMITATIONS OF DATA MINING

Data mining applications can greatly benefit the healthcare industry. However, they are not without limitations[13]. Healthcare data mining can be limited by the accessibility of data, because the raw inputs for data mining often exist in different settings and systems [14], such as administration, clinics, laboratories and more. Hence, the data have to be collected and integrated before data mining can be done. While several authors and researchers have suggested that a data warehouse be built before data mining is attempted, that can be a costly and time-consuming. Secondly, other data problems may arise whereby this include missing, corrupted, inconsistent or non-standardized data such as pieces of information recorded in different formats in different data sources. In particular, the lack of a standard clinical vocabulary is a serious hindrance to data mining applications. Thirdly, there may be ethical, legal and social issues, such as data ownership and privacy issues, related to healthcare data. Fourthly, the successful application of data mining requires knowledge of the domain areas as well as in data mining methodology and tools. Without a sufficient knowledge of data mining, the user may not be aware or be able to avoid the pitfalls of data mining.

VI. SUGGESTIONS

As it obvious that data mining is very important for healthcare and it can improve the situation from health centers to customers, we recommend that the government has to look at the technology, to invest on it, as we believe that it can prove great positive feedback to the lives of residents and organizations too. However, we recommend during implementation, government has to ensure the integration of data, text and digital diagnostic images for images such as X-rays, MRI's whereby by doing so will provide significant help for most residents who cannot afford to travel far for experts to read their images.

VII. CONCLUSION

Currently, Arusha region has few medical personnel and hospitals where data mining can be used as a substitute. We have suggested the implementation and use of data mining technology at Arusha region as this will improve the healthcare industry and well- being of the residents. In the paper we have proposed the integration of data and text mining, using digital diagnostic images which can be brought into healthcare data mining applications. It is our belief that the paper will be a contribution to the data mining and healthcare literature and practice. It also is hoped that this paper can help all parties involved in healthcare reap the benefits of healthcare data mining.

REFERENCES

- [1] T.-H. Chen and C.-W. Chen, "Application of data mining to the spatial heterogeneity of foreclosed mortgages," *Expert Systems with Applications*, vol. 37, pp. 993-997, 2010.
- [2] N. Padhy, "THE SURVEY OF DATA MINING APPLICATIONS AND FEATURE SCOPE Pragnyaban Mishra, Neelamadhab Padhy," 2012.
- [3] R. Barker, "Online crisis communication response: A case study of fraudulent banking transactions in South Africa," *Communicatio: South African Journal for Communication Theory and Research*, vol. 37, pp. 118-136, 2011.
- [4] H. C. Koh and G. Tan, "Data mining applications in healthcare," *Journal of Healthcare Information Management—Vol*, vol. 19, p. 65, 2011.
- [5] McVey, S. 'Data Mining: The Brains Behind eCRM'. (2000). from http://www.technologyevaluation.com/Research/ResearchHighlights/BusinessApplications/2000/11/research_notes/prn_TU_BA_SRM_11_06_00_1.asp (17 May, 2004)
- [6] Liao, S.-h. (2003). 'Knowledge Management Technologies and applications Literature review from 1995 to 2002'. *Expert System with Application*, 25, 155-164.
- [7] A. A. Freitas, "Understanding the crucial role of attribute interaction in data mining," *Artificial Intelligence Review*, vol. 16, pp. 177-199, 2001.
- [8] J. Wang, B. Zhou, and R. Yan, "Benefits and Barriers in Mining the Healthcare Industry Data," *International Journal of Strategic Decision Sciences (IJSDS)*, vol. 3, pp. 51-67, 2012.
- [9] F. Buttle, *Customer relationship management*: Routledge, 2012.
- [10] K. J. Cios, W. Pedrycz, and R. Swiniarsk, "Data mining methods for knowledge discovery," *Neural Networks, IEEE Transactions on*, vol. 9, pp. 1533-1534, 1998.
- [11] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*: Morgan kaufmann, 2006.
- [12] H. Mannila, "Methods and problems in data mining," in *Database Theory—ICDT'97*, ed: Springer, 1997, pp. 41-55.
- [13] Silver, M. Sakata, T. Su, H.C. Herman, C. Dolins, S.B. & O'Shea, M.J. (2001). Case study: how to apply data mining techniques in a healthcare data warehouse. *Journal of Healthcare Information Management*, 15(2), 155-164.
- [14] Benko, A. & Wilson, B. (2003). Online decision support gives plans an edge. *Managed Healthcare Executive*, 13(5), 20.
- [15] Gillespie, G. (2000). There's gold in them thar' databases. *Health Data Management*, 8(11), 40-52.
- [16] Kolar, H.R. (2001). Caring for healthcare. *Health Management Technology*, 22(4), 46-47.
- [17] Relles, D. Ridgeway, G. & Carter, G. (2002). Data mining and the implementation of a prospective payment system for inpatient rehabilitation. *Health Services & Outcomes Research Methodology*, 3(3-4), 247-266.

FPGA Implementation of DA Algritm for Fir Filter

¹Solmanraju Putta, ²J Kishore, ³P. Suresh

¹M.Tech student, Assoc. Prof., Professor Dept of ECE, Sunflower college og Engg. & Technology
Challapalli, Krishna (D), A.P - 521131

ABSTRACT

MAC is composed of an adder, multiplier and an accumulator. Usually adders implemented are Carry- Select or Carry-Save adders, as speed is of utmost importance in DSP. One implementation of the multiplier could be as a parallel array multiplier. The inputs for the MAC are to be fetched from memory location and fed to the multiplier block of the MAC, which will perform multiplication and give the result to adder which will accumulate the result and then will store the result into a memory location. This entire process is to be achieved in a single clock cycle. The architecture of the MAC unit which had been designed in this work consists of one 16 bit register, one 16-bit Modified Booth Multiplier, 32-bit accumulator. To multiply the values of A and B, Modified Booth multiplier is used instead of conventional multiplier because Modified Booth multiplier can increase the MAC unit design speed and reduce multiplication complexity. SPST Adder is used for the addition of partial products and a register is used for accumulation. The product of $A_i \times B_i$ is always fed back into the 32-bit accumulator and then added again with the next product $A_i \times B_i$. This MAC unit is capable of multiplying and adding with previous product consecutively up to as many as times.

I. INTRODUCTION

Due to the intensive use of FIR filters in video and communication systems, high performance in speed, area and power consumption is demanded. Basically, digital filters are used to modify the characteristic of signals in time and frequency domain and have been recognized as primary digital signal processing. In DSP, the design methods were mainly focused in multiplier-based architectures to implement the multiply-and-Accumulate (MAC) blocks that constitute the central piece in FIR filters and several functions. The FIR digital filter

is presented as:

$$y[n] = \sum_{k=0}^{N-1} c[k]x[n-k]$$

Where $y[n]$ is the FIR filter output, $x[n-k]$ is input data and $c[k]$ represents the filter coefficients Equation(1) shows that multiplier-based filter implementations may become highly expensive in terms of area and speed. This issue has been partially solved with the new generation of low-cost FPGAs that have embedded DSP blocks. The advantages of the FPGA approach to digital filter implementation include higher sampling rates than are available from traditional DSP chips, lower costs than an ASIC for moderate volume applications, and more flexibility than the alternate approaches.

II. DISTRIBUTED ARITHMETIC (DA)

Distributed arithmetic (DA) is an important FPGA technology. It is extensively used in computing the sum of products

$$y[n] = \langle c, x \rangle = \sum_{n=0}^{N-1} c[n]x[n]$$

DA system, assumes that the variable $x[n]$ is represented by-

$$x[n] = \sum_{b=0}^{B-1} x_b[n] \times 2^b, x_b[n] \in [0, 1]$$

If $c[n]$ are the known coefficients of the FIR filter, then output of FIR filter in bit level form is:

$$y = \sum_{n=0}^{N-1} c[n] \times \sum_{b=0}^{B-1} x_b[n] \times 2^b$$

In distributed arithmetic form

$$y = \sum_{b=0}^{B-1} 2^b \times \sum_{n=0}^{N-1} f(c[n], x_b[n])$$

In Equation (5.5) second summation term realizing as one LUT. The use of this LUT or ROM eliminates the multipliers [9]. For signed 2's complement number output of FIR filter can be computed as-

$$y = -2^B \times f(c[n], x_B[n]) + \sum_{b=0}^{B-1} 2^b \times \sum_{n=0}^{N-1} f(c[n], x_b[n])$$

Where B represents the total number of bits used. Fig 4.1 shows the Distributed architecture for FIR filter and different with the MAC architecture. When $x[n] < 0$, Binary representation of the input is [10],

$$x[n] = -x_b[0] + \sum_{b=1}^{B-1} x_b[n] 2^{-b}$$

The output in distributed arithmetic form-

$$y = -\left(\sum_{n=0}^{N-1} c[n]x_b[0]\right) + \sum_{b=1}^{B-1} \left(\sum_{n=0}^{N-1} c[n]x_b[n]\right) 2^{-b}$$

If the number of coefficients N is too large to implement the full word with a single LUT (Input LUT bit width = number of coefficients), then partial tables can be and add the results as shown in Fig 4.2. If pipeline registers are also added, then this modification will not reduce the speed, but can dramatically reduce the size of the design

III. IMPLEMENTATION

High Level Specifications are nothing but the requirements to understand and begin the design. In this stage the designer main aim is to capture the behavior of the design using mostly behavioral constructs of the HDL's. The next step after capturing the designs functionality is to segregate the design in all possible ways and try to write a synthesizable code which infers available primitives from the library. Here mostly the mixed style of modeling is used and only synthesizable constructs of the HDL is used. Then comes the synthesis step which is actually target driven. Here we have an FPGA as the target device. Then Implementation is nothing but the process of placing and routing the design on an FPGA. Mostly it is a tool driven and no manual intervention of the designer is required. Designer only needs to specify constrain file in design if any. Below Flow chart shows this in brief.

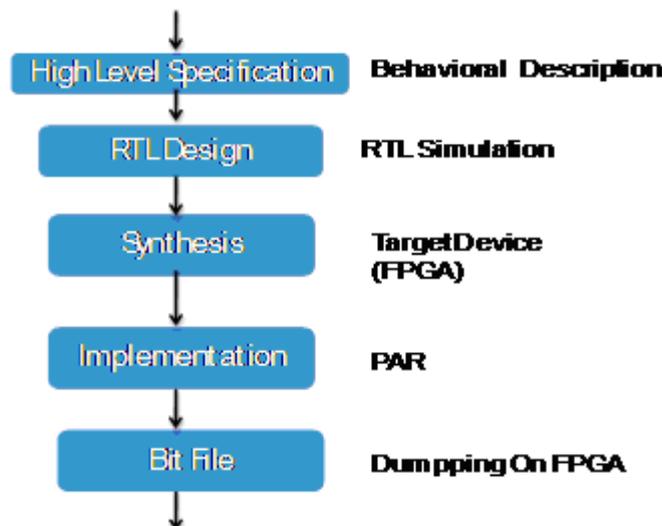


Fig: Design Flow Chart

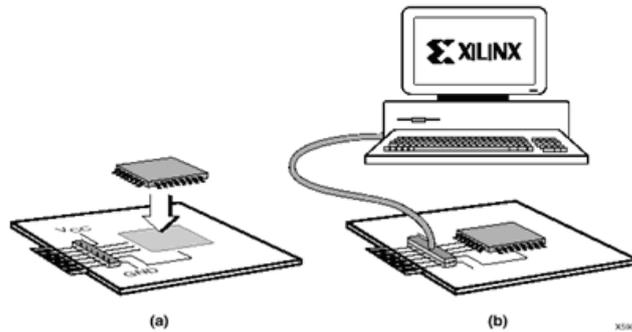


Fig: Bit File Burning

The Final stage after doing the place and route successfully and fully satisfied with mapping and other reports is the bit file generation phase. This bit file has to be downloaded on to FPGA via a JTAG cable. Above Figure shows a simple setup for this.

IV. TARGET DEVICE

Xilinx Spartan3E XC3s100E Device

The below Figure shows the Board with all the peripherals connected to it. The FPGA used belongs to Spartan 3E family and the device is XC3s100E.



Fig: Spartan kit

V. PROGRAMMING

After successfully compiling an FPGA design using the Xilinx development software, the design can be downloaded using the impact programming software and the USB cable. To begin programming, connect the USB cable to the starter kit board and apply power to the board. Then, double-click Configure Device (impact) from within Project. As shown in the figure

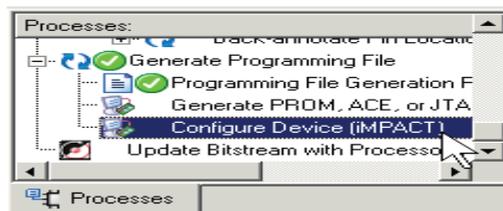


Fig: Configure Device.

If the board is connected properly, the impact programming software automatically recognizes the three devices in the JTAG programming file, as shown in Figure 5.7. If not already prompted, click the first device in the chain, the Spartan-3E FPGA, to highlight it. Right-click the FPGA and select assign New Configuration File. Select the desired FPGA configuration file and click OK. As shown in the fig 5.7 If the original FPGA configuration file used the default Startup clock source, CCLK,IMPACT issues the warning message shown in Figure 5.8. This message can be safely ignored. When downloading via JTAG, the IMPACT software must change the Startup clock source to use the TCK JTAG clock source.

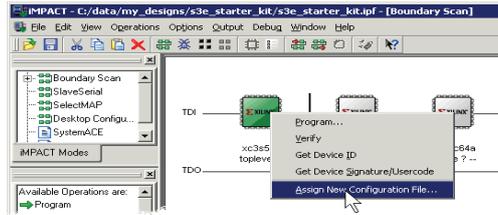


Fig: Assigning configuration File

In above figure we need to assign the correct bit stream file generated to the device and then locate the path of the bit file in the system and then check the box.

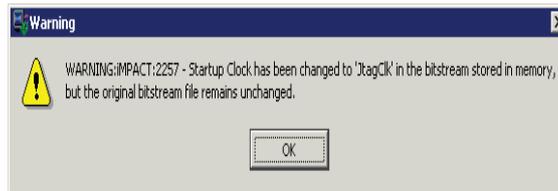


Fig: IMPACT Issues a Warning if the Startup Clock Was Not CCLK
This warning is ignored and checked to continue the burning of bit file to the FPGA.

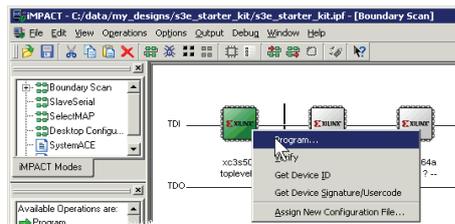


Fig: Program FPGA

When the FPGA successfully programs, the iMPACT software indicates success, as shown in below figure. The FPGA application is now executing on the board and the DONE pin LED lights up.

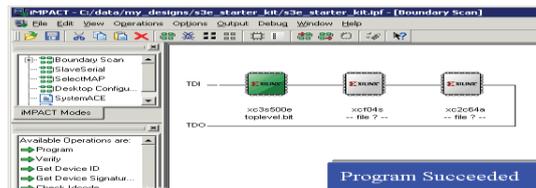
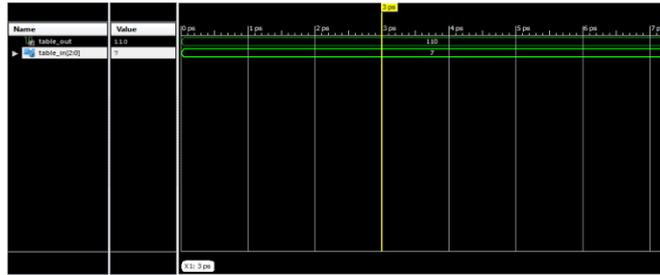


Fig : Successfully Programmed

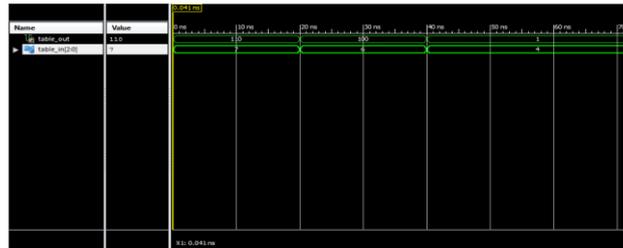
VI. SIMULATION Iteration 1

Name	Value	999,994 ps	999,995 ps	999,996 ps	999,997 ps	999,998 ps	999,999 ps	1,000,000 ps	1,000,001 ps
in[0]	0								
in[1]	011				001				
in[2]	011				011				
in[3]	111				111				
in[4]	11				11				

Iteration 2



Iteration 3



WALLACE TREE

Device	On-Chip Power (W)	Used	Available	Utilization (%)	Supply Summary	Total	Dynamic	Quiescent
Family: Spartan6	Logic: 0.000	67	2400	3	Source: Voltage	Current (A)	Current (A)	Current (A)
Part: xc6slx4	Signal: 0.000	107	---	---	Vccint: 1.000	0.000	0.001	0.000
Package: xqg25	Pin: 0.007	32	132	24	Vccaux: 2.500	0.000	0.001	0.000
Grade: C-Grade	Package: 0.005	---	---	---	Vccqz: 2.500	0.010	0.010	0.000
Revision: Tantal	Total: 0.005	---	---	---				
Speed Grade: -1								
Environment:								
Ambient Temp (C): 25.0	Effective TjA	Max Ambient	Junction Temp		Supply Power (W)	Total	Dynamic	Quiescent
Use counter TjA: No	(C/W)	(C)	(C)					
Custom TjA (C/W): NA	30.6	83.9	26.1					
Initial LPM: D								
Characterization:								
ADVANCED v11.2010.02.23								

DA ALGORITHM

Device	On-Chip Power (W)	Used	Available	Utilization (%)	Supply Summary	Total	Dynamic	Quiescent
Family: Spartan6	Logic: 0.000	22	2400	1	Source: Voltage	Current (A)	Current (A)	Current (A)
Part: xc6slx4	Signal: 0.000	35	---	---	Vccint: 1.000	0.000	0.000	0.000
Package: xqg25	Pin: 0.001	16	132	12	Vccaux: 2.500	0.000	0.000	0.000
Grade: C-Grade	Package: 0.005	---	---	---	Vccqz: 2.500	0.000	0.000	0.000
Revision: Tantal	Total: 0.010	---	---	---				
Speed Grade: -1								
Environment:								
Ambient Temp (C): 25.0	Effective TjA	Max Ambient	Junction Temp		Supply Power (W)	Total	Dynamic	Quiescent
Use counter TjA: No	(C/W)	(C)	(C)					
Custom TjA (C/W): NA	30.6	84.7	25.3					
Initial LPM: D								
Characterization:								
ADVANCED v11.2010.02.23								

VII. CONCLUSION

DA algorithm which is implemented consumes low power of 0.10 watts when compared with recent implementation like Wallace tree which consumes 0.30 watts power .An achievement of 0.20 watts power has been implanted using distributed algorithm. The proposed method has been implemented for 3 bit multiplier and results obtained without any computation error.In general, a multiplier uses Booth’s algorithm and array of full adders (FAs), or Wallace tree instead of the array of FA’s., i.e., this multiplier mainly consists of the three parts: Booth encoder, a tree to compress the partial products such as Wallace tree, and final adder. Because Wallace tree is to add the partial products from encoder as parallel as possible, its operation time is proportional to, where is the number of inputs. It uses the fact that counting the number of 1’s among the inputs reduces the number of outputs into. In real implementation, many counters are used to reduce the number of outputs in each pipeline step.

REFERANCE

- [1] Fast Multiplication Algorithms and Implementation
- [2] Implementation of High Speed FIR Filter using Serial and Parallel Distributed Arithmetic Algorithm
- [3] FPGA Implementation of Digital FIR Filter
- [4] L. Zhao, W. H. Bi, F. Liu, "Design of digital FIR band pass filter using distributed algorithm based on FPGA," *Electronic Measurement Technology*, 2007, vol. 30
- [5] P. Girard, O. Héron, S. Pravossoudovitch, and M. Renovell, "Delay Fault Testing of Look-Up Tables in SRAM-Based FPGAs," *Journal of Electronic Testing*, 2005, vol. 21
- [6] H. Chen, C. H. Xiong, S. N. Zhong, "FPGA-based efficient programmable polyphase FIR filter," *Journal of Beijing Insitute of Technology*, 2005, vol. 14.
- [7] Y. T. Xu, C. G. Wang, J. L. Wang, "Hardware Implementation of FIR Filter Based on DA Algorithm," *Journal of PLA University of Science and Technology*, 2003, vol. 4
- [8] D. Wu, Y. H. Wang, H. Z. Lu, "Distributed Arithmetic and its Implementation in FPGA," *Journal of National University of Defense Technology*, 2000, vol. 22
- [9] L. Wei, R. J. Yang, X. T. Cui, "Design of FIR filter based on distributed arithmetic and its FPGA implementation," *Chinese Journal of Scientific Instrument*, 2008, vol. 29
- [10] W. Zhu, G. M. Zhang, Z. M. Zhang, "Design of FIR Filter Based on Distributed Algorithm with Parallel Structure," *Journal of Electronic Measurement and Instrument*, 2007, vol. 21
- [11] W. Wang, M. N. S. Swamy, M. O. Ahmad, "Novel Design and FPGA Implementation of DAFIR Filters," *Journal of Systems and Computers*, 2004, vol. 13
- [12] M. Nagamatsu et al., "A 15ns 32 x 32-bit CMOS Multiplier an Improved Parallel Structure," *Proc. CICC*, pp.10.3.11989.

A Review on Replica Node Detection Algorithms in Wsns

C. Geetha

Department of CSE, R.M.K. Engineering College,

ABSTRACT

A wireless sensor network is a collection of nodes organized in to a cooperative network. This network is prone to various attacks due to poor security. One of the attacks is clone attack in which an adversary captures some nodes and replicates them including the cryptographic in formation and deploys them in the network. Several algorithms are developed to detect clone attacks, in static WSNs and mobile WSNs. Each one has its own advantages and disadvantages. This paper surveys these algorithms and compares their performance based on parameters like communication cost and memory.

KEYWORDS: attacks, base station, clone attack, communication cost, location, sensor network, witness node.

I. INTRODUCTION

Wireless sensor network, a network of sensor nodes, which are tiny with limited resources that communicate with each other to achieve a goal, through the wireless channels. This network is mainly used in military applications for monitoring security and in civil applications. This network is deployed in harsh and hostile environments. Based on operating nature, it is unattended and prone to various attacks. The basic security requirements of wireless sensor network are integrity, availability, confidentiality and communication. Attacks in wireless sensor networks are classified into internal attacks and external attacks. In internal attacks, compromised nodes can steal secrets from encrypted data, can report wrong information, can report other nodes as compromised nodes and can breach routing by introducing many routing attacks. In external attacks, attackers can capture sensor nodes and reprogram them and can deploy nodes with larger computing resources such as laptops to attack sensor nodes. Several attacks includes Denial of Service, attacks on information, Sybil attack, black hole attack, warm hole attack and clone attack[6][9]. One of the common attacks is clone attack or node replication attack, where an adversary node captures some nodes and makes duplicates of the original node including all cryptographic information and thus inserts these duplicates in the network. These duplicates use the same ID as the original node in the network.

Thus it takes full control over the network. The consequence of this attack is injecting false data, modify the data, initiating a warm-whole attack, dropping packets, thus all these results in leaking of authorized data to an adversary. The simplest way of protecting clone attacks by an adversary node is that, extracts the secret key elements from an attacked node by using a technique called virtue of tamper-resistance hardware. But to implement this technique, the hardware based measures are too expensive in practical. Several algorithms were developed so far to detect clone attacks in both static and mobile sensor networks. The major requirements of all these algorithms are the witnesses and the communication overhead[1].

In this paper we do a study on these algorithms and analyze the performance in terms of detection ratio, speed, communication overhead (memory occupation) and energy consumption. The remaining part of this paper is organized as follows: Section 2 describes the centralized algorithms and distributed algorithms, section 3 compares all these algorithms in terms of Communication overhead and memory usage and section 4 concludes the analysis.

II. DETECTION ALGORITHMS

Based on the detection methodologies, we classify the clone attack detection algorithms as [8][11]

1. Centralized algorithms for static WSNs
2. Centralized algorithms for mobile WSNs

3. Distributed algorithms for static WSNs
4. Distributed algorithms for mobile WSNs

Centralized clone detection algorithms relies on centralized node, may be a base station, where the location or information of all nodes were maintained. But this centralized approach is prone to single point of failure and communication overhead. Only the BS is involved in the detection of clones. In Distributed clone detection algorithms, several nodes are involved in detecting clones. They distribute the location claims to several nodes which are randomly selected and called as witnesses.

A. Centralized Algorithms for Static WSNs

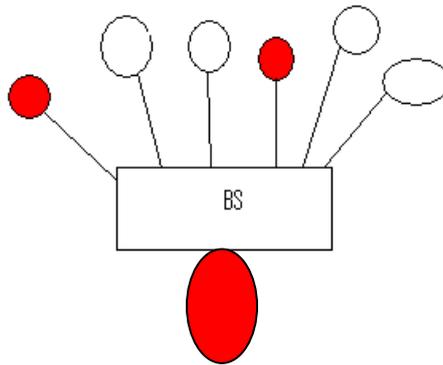
Preliminary Approach: All the nodes send its neighbor's estimated location information and IDS to the base station. The BS compares the IDs and if two nodes with same ID and different location are received by BS, then it finds that there is a cloned node.

Disadvantage

- If the BS fails due to some hardware problem then it is difficult to detect the attacked clone.
- Less number of chances to detect the clone attacks.

Advantage

- The BS contains the ID and information location of all nodes in it.



Nodes with same id and different locations

Fig. 1 Clone attack

Random Key Predistribution: Each and every node is assigned a key to authenticate the nodes. Based on how often these keys are used by the nodes, the key is identified as cloned or not. Each node uses a Bloom filter to count the number of times the key is used. This filter is transferred to the BS, which will verify the count with the predefined threshold, the particular node is found to be a cloned node. But exactly this scheme finds the clone key not clone node. Disadvantage: High false negative and positive rates.

SET detection: This is a method used widely, to detect the clones by using SET operations such as UNION, INTERSECTION of subset of Ids in a network. An intersection of two different subsets must always be empty. If there is a non-empty intersection then the BS revokes that the corresponding two nodes has been cloned.

All previous schemes are used in static WSNs. These schemes are not suitable for mobile WSNs. Because, in mobile WSNs, nodes change the locations often. There is a fact that mobile nodes can't move faster than the maximum configured speed of the network. If any node moves faster than the maximum speed, there must be more nodes with same identity available in the network.

SPRT: When the sensor node moves to a new location, the neighbors ask for the location claim and time information that will be sent to the BS[13]. The BS determines the speed from two consecutive claims and if the speed exceeds the maximum configured speed of network, it found that the mobile node has been replicated. It is the best mechanism in terms of number of observations to reach the decision process. A method in centralized using SPRT, probably a type of SPRT called BIASED-SPRT[15]. SPRT based node compromise detection and revocation schema will not work fast and accurately if more than 50% of the nodes in each zone are

compromised under reasonable configuration. So in order to enhance the SPRT based scheme against the false zone-trust report attack with a large no of compromised nodes the same introduced in the biased-SPRT. In this Biased-SPRT sampling strategy is modified. In that the SPRT takes the sampling leads to acceptance of the H_0 (high trust sample) with the less weight than the ones leading to acceptance of H_1 (low trust samples) ensuring that the false positive rate remains below the desired rate. This modification is called as biased sampling and the corresponding scheme is called as Biased-SPRT.

High trust sample is less likely to be accepted than a low-trust sample if the zone is in un-trust worthy. In accepting the null hypothesis and greater false positive rates biased sampling results in greater delay, but while designing the system these are not major costs. Even if the compromised node is more than 50% then also the biased sampling improves the resilience of the proposed scheme against the false zone-trust report attack.

Performance analysis : The same communication, computation and storage overheads required for biased-SPRT is same as the SPRT because it only changes the sampling strategy of the SPRT and thus does not affect these overheads. However, the change in the sampling strategy affects the average number of samples and attestation overhead of the SPRT.

B. Centralized algorithms for mobile WSNs

Fast Detection – SPRT: SPRT is one dimensional random walk with lower and upper limits. Before random walk, null and alternate hypothesis are defined. Null is associated with lower limit and alternate is associated with upper limit[7]. Each time a mobile sensor node moves to a new location, a signed claim containing its location and time information are send to the neighbors. These neighbors forward the message to BS. The base station computes the speed from every two consecutive claims of a mobile node and performs the SPRT. If the maximum speed is exceeded by the mobile node, set the alternate hypothesis to indicate cloned node.

C. Distributed Algorithms for static WSNs

Using Fingerprint (Real-time Detection): The clones are detected using a fingerprint that includes information of neighboring nodes. Since the fingerprints are fixed on a particular node, it requires additional complex process to add new sensor nodes[10].

Advantages

- When compared to security, this protocol achieves 100% detection of all clones that are attacked assuming that all the messages successfully reach the base station.

Node-to-network delivery: Node to network broadcast is used. Every node collects the IDs and location of its neighboring nodes. When a node receives a broadcast message from the others, the node compares the other neighboring node with its own neighbors, if there is a collision of IDs in the two neighboring nodes of distinct locations, and then the corresponding nodes are cloned and revoked[14].

Advantage

- The distributed methods are more effective than the centralized schemes.
- Failure of the BS node does not cause any problems to the entire system.

Disadvantage:

- High communication overload in the network.

Determined Multicast: To reduce the communication cost, this scheme sends the location information to only a subset of nodes. When a node broadcasts its location claim to all neighbors, and these neighbors in turn send only to limited set of nodes which are called as witnesses. These witnesses are selected based on the function of node ID. If there is a replicated node, any one of this witness may receive the different location claims with same ID and it revokes the replicated node.

Detection using Group Deployment: We define the deployment zone of a group with radius R. Every node discovers its real location. If this node resides outside its home zone, it produces a signed claim by using its private key. Now every node discovers a set of neighbors and asks for authenticated location claim.

If the distance between the node and its neighbor is larger than the communication range then neighbor will be removed from the list. Then it checks the distance between the neighbor and the deployment point is less than R (radius of the group), neighbor is marked as trusted, otherwise mark it as un-trusted[17]. The node forwards regular messages from its un-trusted neighbor, if it has received and verified the location claim, to the

deployment point where all will get the claim message. If any nodes receive conflicting claims, then found that the neighbor node gets replicated.

Randomized Multicast: Same as the previous approach, but the neighbors probabilistically send the location information to randomly selected witnesses. If there is a replicated node, any one of this witness may receive the different location claims with same ID and it revokes the replicated node[4].

Advantage:

- Detects the replication with high probability using relatively limited number of witnesses[16].

Line Selected Multicast: This scheme uses the routing topology to detect the clones. In addition to the witness nodes, the intermediate nodes within the path can check for clones. Each node forwards the claims also saves the claims. For example, a node *a* and clone *a'* in the network. Neighbor of *a* sends the location claim to *r* witnesses. Each node stores this information also. When this information is transferred, on the path any node *w* verifies the signature on the claim and checks for the conflict with the location information on its buffer. If there is a conflict revokes the cloned node. Otherwise store the claim and forwards to the next node[5][12].

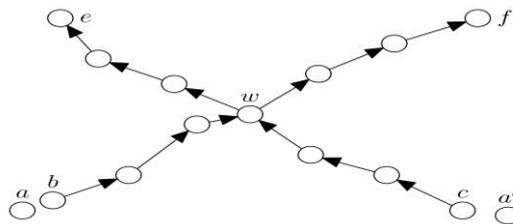


Fig. 2 LSM Approach

Advantage:

- Less communication cost
- High detection rate
- Less storage requirements.

RED: RED is similar, to the RM protocol but with witnesses chosen pseudo-randomly based on a network-wide seed. A random value, *rand*, is shared among all the nodes. This random value can be broadcasted with centralized mechanism. Each node digitally signs and locally broadcasts its claim—ID and geographic location. Each of the neighbors sends (with probability *p*) the claim to a set of $g \geq 1$ pseudo-randomly selected network locations[3].

Agent Based Detection: Every node *A* prepares a signed location claim. The mobile agent gets the signed claim which is visited by it. The nodes information matrix is acquired through mobile agent routing algorithm. Each node *A* gets the information matrix verifies the signature and checks the distance between the neighbors and this cannot be bigger than the transmission range. When more than one entry for signed claim made in a single cell of an information matrix of one node, revokes the procedure for replicas[2].

D. Distributed algorithms for mobile WSNs

XED (eXtremely Efficient Detection): Every sensor node is having a random number generator. When a node encounter another node, they exchange the random numbers. Once again the same nodes meet, they verify the random numbers exchanged already. If no match, clone node found[18].

Advantage:

- Communication cost is constant.
- Location information is not required to detect the clone node.

Disadvantages:

- Vulnerable to smart attackers.

SDD: If a node *a* does not meet the node *b* within twice the specified interval Δ , then the node is removed from the network and clone node is found[19].

Advantages:

- Each node is considered as witness.

Disadvantages:

- High communication cost and Less detection ratio.

EDD (Efficient and Distributed Detection): The algorithm has two phases. In the first phase, the interval Δ is calculated and in the second phase, the messages are exchanged to find the clone[20].
Disadvantages: Vulnerable to smart attackers

SEDD (Storage EDD): A node monitors only a set of nodes and exchanges the messages within the set[20].
Advantages: Reduces the number of messages exchanged.

UTSLE (Unary-Time-Location Storage & Exchange): This does not require any routing messages. It uses the time-location claim. When two witnesses encounter each other, the time-location claims are checked, which are stored only once.

MTLSD (Multi-Time-Location Storage & Diffusion): Stores more time-location claims for each tracked node. It provides excellent resiliency and sub-optimal detection probability with modest communication overhead.

III. COMPARISON

TABLE I
COMMUNICATION COST AND MEMORY

Protocol Name	Communication Cost	Memory
Preliminary Approach	$O(n)$	$O(n)$
SET	$O(n)$	$O(d)$
SPRT	$O(n)$	$O(d)$
Deterministic Multicast	$O(g \ln g \sqrt{n} / d)$	$O(g)$
RED	$O(r \sqrt{n})$	$O(r)$
Randomized Multicast	$O(n^2)$	$O(\sqrt{n})$
Line-Selected Multicast(LSM)	$O(n \sqrt{n})$	$O(\sqrt{n})$
SDC	$O(r_f \sqrt{n}) + O(s)$	g
P-MPC	$O(r_f \sqrt{n}) + O(s)$	g
XED	$O(1)$	
EDD	$O(1)/O(n)$	$O(N)$
UTLSE & MTLSD	$O(n)$	$O(\sqrt{n})$

Where,

n - No .of nodes in the network, d - Degree of neighboring nodes, r - Communication radius, g -no. of witness nodes, r_f - no. of neighbor nodes forwarding claims.

IV. CONCLUSION

In this paper a study of various clone detection approaches was done. The distributed approach is more efficient than the centralized one because of single point of failure. The selection of witnesses is different for the approaches discussed for static WSNs. When the mobility is added, the algorithms become more complex. In future, we had a plan to design a new approach based on RED and LSM protocols, which might meet the requirements of clone detection algorithms and also high detection ratio with less time and communication cost.

REFERENCES

- [1] Mauro Conti, Roberto Di Pietro , Luigi V. Mancini, and Alessandro Mei, "Requirements and Open Issues in Distributed Detection of Node Identity Replicas in WSN", IEEE International Conference on Systems, Man and Cybernetics, 2006, Taiwan.
- [2] D.Sheela , Srividhya .V.R, Vrushali , Amrithavarshini and Jayashubha J."A Mobile Agent Based Security System of Wireless Sensor Networks against Cloning and Sink Hole Attacks", ICCTAI'2012.
- [3] Mauro Conti, Roberto Di Pietro , Luigi Vincenzo Mancini, and Alessandro Mei, "Distributed Detection of Clone Attacks in Wireless Sensor Networks", IEEE Transactions On Dependable And Secure Computing, VOL. 8, NO. 5, 2011.
- [4] Bryan Parno, Adrian Perrig, Virgil Gligor , "Distributed Detection of Node Replication Attacks in Sensor Networks".
- [5] Wen TaoZhu, JianyingZhou, RobertH. Deng, FengBao , "Detecting node replication attacks in wireless sensor networks: A survey", Journal of Network and Computer Applications 35 (2012) 1022–1034.
- [6] Dr. G. Padmavathi , D.Shanmugapriya "A Survey of Attacks, Security Mechanisms and Challenges in Wireless Sensor Networks " International Journal of Computer Science and Information Security, Vol.4 No. 1&2,2009.
- [7] Jun-Won Ho, Matthew Wright, and Sajil K. Das "Fast Detection of Replica Node Attacks in Mobile Sensor Networks Using Sequential Analysis". IEEE Transactions on Mobile Computing June 2011 (vol. 10 no. 6) pp. 767-782
- [8] Kwantae Cho, Minbo Jo, Member , IEEE, Tackyoung Kwon, Hsiao - Hwa Chen, Fellow, IEEE ,and Dong Hoon Lee, Member, IEEE" Classification and Experimental Ana,y-sis for Clone Detection Approaches in Wireless Sensor Networks" IEEE SYSTEM JOURNAL .
- [9] TEODOR-GRIGORE LUPU , Vasile Parvan 2,300223, Timisoara, " Main Types of Attacks in Wireless Sensor Networks " Recent Advances in Signals and Systems.
- [10] Kai Xing, Fang Liu Xiuzhen Cheng, David H. C .Du, "Real-time Detection of Clone Attacks in Wireless Sensor Networks" The 28th International Conference on Distributed Computing Systems.
- [11] V. Manjula, C. Chellappan "Replication Attack Mitigation For Static And Mobile Wsn" International Journal; of Network Security&Its Application(IJNSA)Vol.3,No.2, March 2011.
- [12] Bryan Parno, Adrian Perrig, Virgil Gligor , "Distributed Detection of Node Replication Attacks in Sensor Networks". Proceeding of 2005 IEEE Symposium on Security and Privacy Pages 49-63
- [13] Jun-Won Ho, Matthew Wright, Member, IEEE, and Sajil K. Das , Senior Member, IEEE," Fast Detection of Mobile Replica Node Attacks in Wireless Sensor Networks Using Sequential Hypothesis Testing" IEEE TRANSACTIONS ON MOBILE COMPUTING, VOL.10, NO.6,June 2011.
- [14] V. Ram Prabha, P . Latha, "An Overview of Replica Node Detection Wireless Sensor Networks" International Conference on Recent Trends in Computational Methods, Communication and Controls(ICON3C 2012)Proceedings Published in International Journal of Computer Applications(IJCA).
- [15] Jun- Won Ho, Member, IEEE Computer Society, Matthew Wright, Member, IEEE, and Sajal K. Das, Senior Member, IEEE" Zone-Trust :Fast Zone-Based Node Compromise Detection and Revocation in Wireless Sensor Networks Using Sequential Hypothesis Testing" IEEE TRANSACTIONS ON DEPENDABLE ANDSECURECOMPUTING,VOL.9,NO.4,JULY/AUGUST 2012.
- [16] Tamara Bonaci, Phillip Lee, Linda Bushnell, Radha Poovendran" A convex optimization approach for clone detection in wireless sensor networks" Contents List available at Elseivere Science Direct.
- [17] Jun-Won Ho, Matthew Wright, Donggang Liu, and Sajil K.Das , " Distributed detection of replica node attacks with group deployment knowledge in wireless sensor networks " Ad Hoc Networks 7(2009)-1476-1488.
- [18] Mauro Conti et al," Emergent Properties: Detection of the Node-capture Attack in Mobile Wireless Sensor Networks" WiSec'08, March 31–April 2, 2008, Alexandria, Virginia, USA.
- [19] Chia-Mu Yu et al," Efficient and Distributed Detection of Node Replication Attacks in Mobile Sensor Networks" 2009 IEEE
- [20] Chia-Mu Yu et al," Mobile Sensor Network Resilient against Node Replication Attacks" 2008 IEEE