

# MEDICAL DATA MINING: A REVIEW

Amrita Kundu<sup>1</sup>, Pallavi<sup>2</sup>

<sup>1,2</sup> Department of Computer Science, Banasthali University, (India)

## ABSTRACT

*Data Mining refers to the mining of useful and interesting patterns from large data sets. Since its advent in the early 1980s data mining has made remarkable progress because of its use in industry, medical science, scientific applications, web etc. Medical data now a day is available in abundance but without proper mining they cannot be used. Using data mining techniques on medical data several critical issues can be understood better and dealt with starting from studying risk factors of several diseases to identification of the diseases occurring frequently or taking care of hospital information systems. In this paper light has been thrown on various data mining approaches for efficient management of medical data that can serve the mankind which is the prime motto behind any research work.*

**Keywords:** *Data Mining, Decision trees, KDD, Medical data, Naïve Bayes.*

## I INTRODUCTION

The goal of data mining is to learn from data [4]. The strategy used may vary as per the requirement. Data mining is an interdisciplinary field and is gaining popularity because of exploring Database technology, Information Science, Machine learning and Neural networks along with the Statistical techniques. Though data mining algorithms are not applied on the medical data by common people but the knowledge obtained can be very useful for them if shared with in an understandable form. Some of the applications of data mining on medical data include classification of several medical images like X-ray images or MRI images can serve for better diagnosis of any abnormality in the body, clustering the patient records any chronic disease to obtain the knowledge about the spread of the disease, analyzing data in healthcare or survival chances of a patient suffering from diseases like cancer.

## II LITERATURE REVIEW

**2.1** In year 2004, Mary K. Obenshain [1] applied data mining in three healthcare arena namely hospital infection control, ranking hospitals and identifying high-risk patients.

1. Hospital infection control- Infectious diseases breakout at different times in different geographic areas and there has been a rapid increase in the number of drug-resistant infections too. A surveillance system has been discussed that uses association rule on culture and patient care data obtained from laboratory information systems to generate useful patterns monthly that can be used by an expert taking care of infection control [1].

2. Ranking hospitals- Mining on the reports from the healthcare providers have been discussed to rank hospitals and healthcare plans. According to the reports provided the hospitals should be ranked according to their standard by the organizations [1].

3. Identifying high-risk patients- A robust data mining and model-building solution has been suggested for identifying the patients tending towards high-risk conditions so that the quality of healthcare of patients can be improved [1].

*Critic:* The role of association rule is to segregate the uninteresting patterns from the interesting ones using a support-confidence framework. For ranking hospitals it is necessary that the hospitals provide data correctly else patient mortality rate cannot be predicted accurately. Identifying high-risk patients may help decrease their number in future.

**2.2** In year 2005, authors Md. Rafiqul Islam et.al. [2]made an attempt to implement data mining in image archiving systems using an algorithm based on an inductive decision tree to learn the attributes of lung cancer. Tomograms of 250 patients were considered for the experiment from which patients with pulmonary nodules of size up to 5 cm were chosen . Before application of the algorithm the images were preprocessed and feature subset was selected for enhancing the accuracy of experiment and reducing unnecessary data.

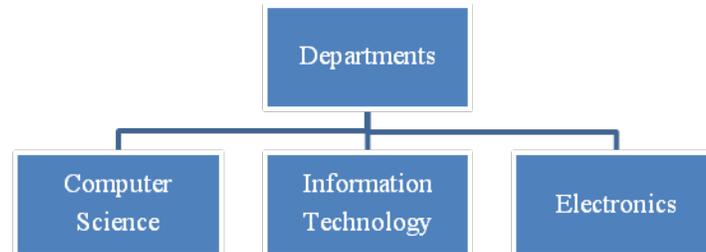
*Critic:* There is abundance of image data but only applying proper data mining techniques can help extract the attributes necessary to study any disease with accuracy.

**2.3** In year 2006, authors Abdelghani Bellaachia et.al have given their study [3]. Breast cancer is a common form of cancer now a day. The chances of survivability of cancer patients vary from case to case and the stage to stage. The authors used the Naïve Bayes, the back-propagated neural network and the C4.5 decision tree algorithm to predict that and compared the three.

**TABLE-1 Classification techniques used with their accuracy**

CLASSIFICATION TECHNIQUE	ACCURACY (%)
Naïve Bayes	84.5
Neural Net	86.5
C4.5	86.7

*Critic:* A decision tree is similar to flow chart having a tree structure where tests are performed at each level except the last one which gives the result. It is a powerful means for classifying a data set. Predicting the survival time of a cancer patient by this means can actually help the doctor to plan further treatment procedures and the patient to take care of himself so that the time can be extended as much as possible.



**Fig. 1. A decision tree**

**2.4** In year 2007, Ghim-Eng Yap et. al.[4] had given their study:

a) Whenever we deal with data sets errors are evident. These errors may have considerable effect on the output of results when not diagnosed and this effect may be very serious when it's about the medical diagnosis of a disease. In this paper the authors made an attempt that deal with this issue using a knowledge discovery approach by Bayesian network learning [4]. Also a novel procedure for handling the error is implemented to deal with the uncertainty caused by them. This ultimately leads to reliable distinction between cancer affected (here ovarian cancer) and normal patients. The use of Bayesian network has been shown in three steps:

1. Presenting serum expression profile to a learned Bayesian network in case of an unseen sample. .
2. Based on it the network is allowed to update the posterior probabilities of all the nodes it has.
3. Concluding if a patient is suffering from cancer or not based on the largest posterior probability.

Before this erroneous markers are discovered and randomness in the behavior of studied protein is eliminated using a threshold value which is considered 0.1.

*Critic:* The biological data if noisy can lead to major problem in diagnosis result while managing this noise can give equally good result in cancer detection even with the use of only ten proteins. Such type of tool discussed can prove to be very efficient in screening the patients.

b) Several diseases have a hairline difference between them and are considered as similar. However distinguishing them is necessary to provide the most appropriate treatment to the patient. In this paper an attempt has been made by the authors to meet this challenge using an open source data mining toolbox. Here the two diseases considered are thrombolic brain stroke and embolic brain stroke as given by Petra Kralj et.al. [5]. The experiment included collecting the CT of total 290 patients belonging to the above mentioned categories and using decision tree induction to obtain the contrast sets. Other than this the physical examination of the patients, their ECG, laboratory and other diagnosis measures were also considered [5]. Then data mining was applied to mine the contrast sets to classify the patients belonging to the two categories. This can be done using the Apriori algorithm. A learning rule is finally applied.

*Critic:* Subgroup discovery approach involves the mining of the interesting groups with the help of a target variable. In this the predictive and descriptive induction are combined. Contrast set mining is used to differentiate between the contrasting groups in a data set. It can be implemented using special algorithm STUCCO or decision tree induction and learning rule[5]. The authors performed contrast set mining through subgroup discovery to differentiate between the patients belonging to the two different classes.

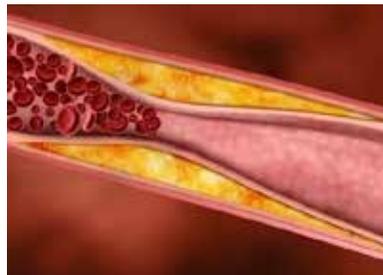
**2.5** In year 2008, Dimitris Bertsimas [6] helped with their comments:

a) The most common problem faced by the patients is not the disease they are suffering from but the expenditure of treatment. The common people are the main victims of this problem. So the authors made an attempt to estimate the medical cost and using the past cost pattern to predict the future cost. The data set used was the medical and pharmaceutical claims data for 838,242 individuals. The diagnostic, procedure and drug related information are present in the claims [6]. Cost bucketing is performed on the samples and the Baseline method is used along with classification trees and clustering to group the members with similar cost characteristics [6].

*Critic:* Care must be taken of the errors for accurate prediction of medical costs. Here the measures of errors used are the hit ratio, the penalty error and the absolute prediction error (APE) [6]. This method can be used by the insurance companies too for pricing the health insurances. Also the comparison with past expenses can be beneficial in predicting the rise in future too.

b) The authors in this paper tried to use sequential data mining to prevent atherosclerosis, a disease characterized by thickening of the arterial walls by deposition of white blood cells. The risk factors of the disease have been taken into account. Case study: STULONG has been used which is specially done to detect the risk factors and other related conditions of atherosclerosis. The windowing approach is used to decompose the data into several disjoint windows. OAR algorithm and AR mining procedure are applied for mining association rules as given according to Jiřr'ı Kl'ema et.al. [7]. Episode rule mining is used to mine the sequential patterns.

*Critic:* Sequential data mining deals with discovering patterns in data that are statically relevant. Sequential data has significant potential in medical science but they are not present in a form suitable for the direct application of general data mining algorithms.



**Fig. 2. Atherosclerosis**

**2.6** In year 2009, the author Ruban D. Canlas Jr [8] explains the use of data mining techniques in medical research and public health. Through data mining fraudulent insurance claim can be detected, better health-policy making health centers can be analyzed. In Health Sector Use of data mining and KDD was introduced in this area by Wilson et al [8]. The importance of data mining in medicine and public health can be summarized as-

1. Data Overload: There is a bulk of computerized health record. The medical breakthroughs have slowed down and complexity of the present-day medical information is high, so data mining is best-suited to discover the knowledge. [8]

2. Evidence-based medicine and prevention of hospital errors: In this we apply data mining on their existing data; it can discover useful and life-saving knowledge. By mining hospital record we reduce the rate of error that have made. [8]

3. Policy-making in public health: Open source java based data mining tool Weka and J48 [8] is used to find out the similarities between the community health centers. Data mining and decision support methods, can lead to better performance in decision making. [8]

4. More value for money and cost savings: Through data mining we extract more knowledge from existing data at minimal cost. Data mining is applied to discover knowledge about fraud in credit cards and insurance claims. [8]

5. Early detection and/or prevention of diseases: Classification algorithm is used for early detection of heart disease and heart related disease is the major public concern in the world. Data mining tool is to aid in monitoring trends in the clinical trials for cancer vaccines. [8]

6. Early detection and management of pandemic diseases and public health policy formulation: Health experts have decided to apply data mining in early detection and management of pandemics diseases. In this we have applied simulation and spatial data mining to find interesting characteristics of disease out-break. In 2005 introduced WSARE [8], an algorithm which is used to detect outbreak in early stages. [8]

7. Non-invasive diagnosis and decision support: Thangvel et al. (2006) decided to use K-means clustering algorithm to analyze the cervical cancer patients. The predictive result concluded from K-means clustering algorithm is better than existing medical opinion. Data mining is used to enhance computer-aid diagnosis and endoscopic ultra sonographic elastography and this create new non-invasive cancer detection [8].

8. Adverse drug events (ADEs): Data mining used in US food and drug administration to discover knowledge about drug side effects in their database. The algorithm which is used is called MGPS (Multi-item Gamma Poisson Shrinker) and it was able to detect 67% of ADEs five years before than they were detecting through traditional approach. [8]

*Critic:* Application of data mining techniques in medical field is a big challenge due to the changing behavior from patient to patient. But still data mining algorithms are used quite successfully to extract better information in case of fraud detection in health policies and as well as in early detection of disease such as heart and pandemics from existing data which is stored in the medical database.

*Critic:* Swarm Intelligence (SI) is the collective behavior of decentralized, self-organized systems natural or artificial introduced by Gerardo Beni and Jing Wang in the year 1989 and is used in artificial intelligence.



**Fig. 3. Swarm Intelligence in fishes.**

Comparison table-

**TABLE-2 Algorithms used in the paper and their significance**

Algorithm Used	Significance
ACO	Works well on nominal and categorical attributes
PSO	Low in Spatial Complexity
Combined ACO/PSO	To find one classification rule at a time

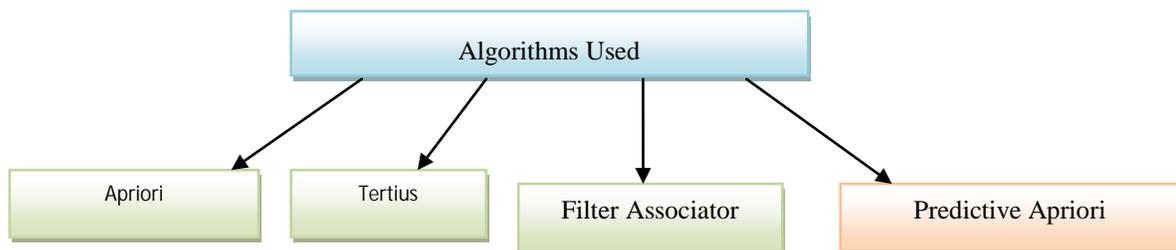
**2.7** In Year 2011, Veenu Mangat, [9] the authors have discussed the use of data mining in medical records for enhancement of strategic decisions. Weka version 3.6.0 is used as software for the data mining analysis. Data set were converted into Weka data format after that different associative algorithm was executed on data set and their results were comparatively analyzed. There are four algorithms in which comparative analysis is done. After comparative analysis it was found that predictive Apriori is best fitted algorithm in strategic decisions with highest and lowest accuracy values being 0.99498 and 0.9733 Yilmaz GOKSEN et.al. [10].

Apriori: It generates association rules by using frequent item sets and it generate longer candidate item sets from shorter ones. It reduces minimum support until it find required number of association rules. [10]

Filter Associator: It passess the data through filter before it reaches to the associator. User can configure both base associator and filter. [10]

Predictive Apriori: To find out best ‘n’ association rule it merges confidence and support into single measure of predictive accuracy. [10]

Tertius: Rules are find according to a confirmation measure. Rules were seeked with multiple conditions like apriori difference is that conditions are applied in OR operations together instead of using and operation [10].



**Fig. 4. Algorithms discussed in the paper.**

*Critic:* We have applied association algorithm to extract association rule, but it is noticed sometimes data sets itself might not be suited for association tasks in data mining. No data mining tools and model will provide 100% accuracy by itself in pure and robust automated system. We need experts and managers for checking the final result.

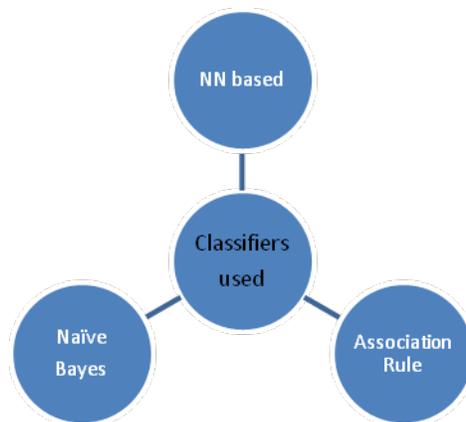
**2.8** In year 2012, Shweta Kharya [11], has discussed some classifiers:

a) In the paper author have explored the applicability of decision trees to find a group with high-susceptibility of suffering from breast cancer. One of the most widely used and practical methods for classification is decision tree learning. To generate best decision tree be have used different parameter like confidence factor, pruning. For anomaly detection, classification and automatically categorize medical images on real mammograms two data mining techniques, association rule mining and neural networks is used. There are certain classifier which is explained in the paper and used for diagnosis and prognosis of cancer disease. [11]

1. Association rule based classifier: In this we find out association rule for the medical data. One basis of association rule classification system is constructed and it will categorize the mammograms as normal, malign or benign. [11]

2. Neural network based classifier system: In neural network we classify the medical data set and train the neural network with breast cancer data base by using feed forward network and back propagation learning algorithm with momentum and learning rate. [11]

3. Naïve Bayes Classifier: Analysis of the prediction of survivability rate of breast cancer is done by using naïve Bayes, back- propagated neural network, and decision tree [11].



**Fig. 5. The classifiers discussed in the paper.**

*Critic:* In diagnosis and prognosis of cancer disease we have used different data mining classifiers and soft computing approaches. Decision tree is found to be best predictor with greater amount of accuracy as compared to other techniques.

b) In this paper the authors have discussed the issue that availability of huge amounts of medical data leads to need for powerful data analysis tools to extract useful knowledge. Five data mining techniques classification by decision tree induction, Bayesian classification, Neural networks, Support Vector Machines (SVM) and classification based

on associations have been analyzed and it was found that decision tree and SVM are most effective for heart disease diagnosis as given by Aqueel Ahmed et.al. [12]. Various classification techniques for data mining are examined and reported that data mining technique shows the 92.1 % - 91.0 % accuracy for the heart diseases. Here usage of various parameters (age, fasting blood sugar, sex etc.) increases the accuracy for the heart disease patient using data mining techniques.

*Critic:* Healthcare system has huge data available but effective analysis of those using proper tools is the job of data mining techniques. Finding out the hidden relationships and patterns in the data can help predicting a disease very accurately.

**2.9** In year 2013, Taranath NL et.al.[13] focused on novel framework of medical decision support system for missing data. In this author have used the machine learning approach, automatic learning is used in certain tasks such as medical imaging. In this there are two tasks to be performed: identification and extract informative sentence on disease , fine grained classification according to semantic relation of the sentence on the basis of diseases and treatment. [13]

*Critic:* In order to offer the required accuracy for this domain both the approaches of integrating machine learning and ontological reasoning is found best approaches. The framework of medical decision support system is emerging application in medical sector.

**b)** The paper contain the idea of developing a prediction model that can predict heart disease cases based on measurements taken from transthoracic echocardiography examination. KDD has been used. Prediction system built with aid of data mining techniques like decision trees, naïve Bayes and neural network [14]. The models used along with the accuracy measures are:

**TABLE-3 Models used in the paper and their accuracy**

Model Used	Accuracy(%)
J48 unpruned with all attributes	94.29
J48 pruned with all attributes	95.41
Naive Bayes with all attributes	91.96
Naive Bayes with selected attributes	92.42

*Critic:* The goal of paper is to develop prediction model, but researcher has planned to perform additional experiments with more dataset and algorithms to improve the classification accuracy. On the basis of this model that can predict specific heart diseases.

**2.10** In year 2014, Matthew Herland et.al. given their review on big data in the field of health informatics[15]:

*a)* Since the amount of data in health informatics is very vast, the authors in this paper discussed the concept of big data analysis for improving the quality of healthcare. Problems regarding analyzing such huge data in a reliable

manner are also discussed. Big data tools and approaches at various levels are dealt with live molecular, tissue, patient and population [15].

**TABLE-4 Summary of studies in paper**

Data level	Question level(s) answered
Molecular	Clinical
Tissue	Human scale biology, Clinical
Patient	Clinical
Population	Epidemic-scale, Clinical

*Critic:* In future work needs to be done on handling vast amounts of gene probe and select the kind of subset that can provide the best correlation. Difficulty is faced while dealing with MRI data that are high resolution data and numerous samples are required to perform the research work of acceptance level with those.

b) Apriori algorithm is a general algorithm in data mining .In the paper , year 2014 the authors Gitanjali J et.al.[16], the authors used this algorithm to find out the frequency diseases using the data set of months, diseases and the instances of their occurrences. Experimental results are shown in the form of bar graphs to show frequency of occurrence of different diseases. The steps used are:

1. Scanning data set D to generate list of candidates [16].
2. Compare candidate support count with min support count [16].
3. Scan D for count of each candidate and compare candidate support count with min support count remaining in list [16].

*Critic:* The frequency of occurrence of diseases varies with geographic area, habits of people, season etc.

### III CONCLUSION

Medical data available is huge and hence analyzing or classifying them with efficiency is a challenge. Also high resolution data like MRI are to be mined in large amount for predicting critical diseases of brain with accuracy. Several classifiers have been implemented till date including artificial neural networks but proper learning technique must be implemented to get the desired results. In the presence of all this challenges also data mining algorithms have shown eye catching results in different domains of medical science but more work is required to be done in managing different variety of diseases, major or minor because mostly minor disease affect people which may turn major if not treated correctly on time.

### REFERENCES

- [1] Mary K. Obenshain, "Application of Data Mining Techniques to Healthcare Data," Statistics for Hospital Epidemiology. North Carolina, vol. 25 No. 8, pp. 690-695, August 2004.
- [2] Md. Rafiqul Islam, Morshed U. Chowdhury and Safwan Mahmood Khan\*, Medical Image Classification Using an Efficient Data Mining Technique. Dhaka, Bangladesh, vol.12, Complexity International, 2005, pp.01-09.

- [3] Abdelghani Bellaachia and Erhan Guven, "Predicting Breast Cancer Survivability Using Data Mining Techniques," Washington DC 20052, The George Washington University, 2006, pp. 01-04.
- [4] Ghim-Eng Yap, Ah-Hwee Tan and Hwee-Hwa Pang, "Learning Casual Models for Noisy Biological Data Mining: An Application to Ovarian Cancer Detection," Singapore 639798. Association for the Advancement of Artificial Intelligence 2007, pp.354-359.
- [5] Petra Kralj1, Nada Lavra-c1;2, Dragan Gamberger3, Antonija Krsta-ci, "Contrast Set Mining for Distinguishing Between Similar Diseases," Slovenia and Croatia, Technology Project "Knowledge Technologies",2007.
- [6] Dimitris Bertsimas, Margrét V. Bjarnadóttir, Michael A. Kane, J. Christian Kryder, Rudra Pandey, Santosh Vempala and Grant Wang, "Algorithmic Prediction of Health-Care Costs," Operations Research, vol.56, No.6, pp. 1832-1892, November-December 2008.
- [7] Jiří Klíma, Lenka Nováková, Filip Karel, Olga Štěpánková, and Filip Zelezný, "Sequential Data Mining: A Comparative Case Study in Development of Atherosclerosis Risk Factor," IEEE Transactions on System, Man and Cybernetics. vol. 38 ,No.1, pp. 01-12, January 2008.
- [8] Ruban D. Canlas Jr., "Data Mining in Healthcare:Current Applications and Issues"MSIT, August 2009
- [9] Veenu Mangat, "Swarm Intelligence based Technique for Rule Mining in the Medical Domain"International Journal of Computer Applications ,vol. 4, no.1, July 2010
- [10] Yilmaz GOKSEN, Mete EMINAGAOGLU and Onur DOGAN, "Data Mining in Medical Records for the Enhancement of Strategic Decisions:A Case study"Scientific Bulletin – Economic Sciences, Vol. 10 / Issue 1,2011
- [11] Shweta Kharya, "Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease",International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.2, April 2012.
- [12] Aqueel Ahmed, Shaikh Abdul Hannan., "Data Mining Techniques to Find Out Heart Diseases,"International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-1, Issue-4, September 2012.
- [13] Taranath NL, Dr. Shantakumar B Patil, Dr. Premajyothi Patil, Dr. C.K. Subbaraya, "Medical Decision Support System for the Missing Data using Data Mining-A Survey"International Conference on Contemporary Computing and Informatics,2014 IEEE.
- [14] Abhishek Taneja, "Heart Disease Prediction System Using Data Mining Techniques" Oriental journal of Computer science & technology, December 2013, Vol. 6, No.4: pp. 457-466.
- [15] Matthew Herland, Taghi M Khoshgoftaar and Randall Wald\*, "A Review of Data Mining Using Big Data in Health Informatics ," Herland et al. Journal of Big Data 2014, 1:2, USA.
- [16] Gitanjali J, C.Ranichandra,M.Pounambal, "APRIORI Algorithm Based Medical Data Mining for Frequent Disease Identification," International Journal of Information Technology (IIJIT), vol.2, issue 4, April 2014.

Year	Technique used	Objective of the paper
2004	Decision trees, Neural Network,	Applications of data mining techniques to health care data.

	Naïve Bayes classification, Logistic Regression	
2005	Decision trees, Neural Network	Classification of medical image.
2006	Weka toolkit, C4.5 decision tree	Predicting Breast cancer survivability.
2007	a) Clustering b) KDD using Data mining tools	a) Analysis of clinical courses of chronic Hepatitis. b) Managing noisy data while ovarian cancer detection.
2008	a) Bucketing, clustering, classification trees b) Trend analysis, windowing	a) Predicting health-care costs. b) Development of Atherosclerosis risk factor.
2009	KDD using Data mining tools	Studying current applications and issues related to healthcare.
2010	Swarm Intelligence algorithms	Rule mining in medical domain.
2011	Estimation, prediction, classification, clustering, association	Mining medical records for the enhancement of strategic decisions.
2012	a) Classification, Neural Network, Association rule mining, C4.5 decision tree, Naïve Bayes. b) Association rule, clustering	a) Diagnosis and Prognosis of cancer disease. b) Finding out heart diseases.
2013	a) Knowledge-based and learning-based system b) Neural network, decision tree, Naïve Bayes	a) Medical decision support system for the medical data. b) Designing heart disease prediction system.
2014	a) Big Data tools b) Apriori algorithm	a) Data mining using big data in health informatics. b) Frequent disease identification.

### Annexure

#### Biographical Notes

**Ms. Amrita Kundu** is presently pursuing M.Tech. first year in Computer Science Engineering Department from Banasthali Vidyapith, India.

**Ms. Pallavi** is presently pursuing M.Tech. first year in Information Technology Engineering Department from Banasthali Vidyapith, India.

# HIGH PERFORMANCE FIR FILTER USING CARRY SELECT ADDER TO REDUCE AREA-DELAY- POWER

Anline Reeta<sup>1</sup>, N. Kaleeswari<sup>2</sup>

<sup>1</sup>Department of Electrical and Electronics, B.E, Ketti (India)

<sup>2</sup>Assistant Professor, Department of Electronics and Communication, ME (India)

## ABSTRACT

Ripple carry implementation allows adder tree to minimize hardware cost, unfortunately it sacrifice timing and gives low speed operation. To outperform this high speed adder is proposed and analyzed for real time speech signal applications. Multiple constant multiplication scheme is the most effective common sub expression sharing technique which is used for implementing the transposed FIR filters. The resource minimization problem in the scheduling of adder tree operation based Mixed integer programming (MIP) algorithm for more efficient MCM based implementation of FIR filters are identified. The proposed adder tree consists of carry select adder to improve the speed of the FIR filter.

**Keywords:** Carry Select Adder Tree Implementation, Multiple Constant Multiplication, Transposed Fir Filter.

## I. INTRODUCTION

Digital signal processing technology and its advancements have dramatically impacted our modern society everywhere. Without DSP, we would not have digital audio and speech, Digital telephone, Automobile industry, Electronic communications, Medical imaging equipment, Multimedia applications. The signals are usually processed in digital representation. , so speech processing can be regarded as a special case of digital signal processing the digital filter is the most important system in speech processing. It is used to reduce the noise in an information bearing signal. Based on the impulse response filters are classified as two.

1. Finite impulse response filter (FIR)
2. Infinite impulse response (IIR).

FIR filter is designed using finite number of impulse response. IIR filter does not provide the stable output, So that a design of an efficient transposed FIR filter which produces stable output is produced. The complexity of the FIR filter is dominated by the multiplication of the input samples with filter coefficients. Filters employ a large number of multipliers that lead to excessive area and power consumption. But the filter coefficients are constant for a given filter, so that multiplications are implemented by a network of adders and sub tractors. Where the number of additions and subtractions are minimized by a constant multiplication scheme. In the transposed fir filter, the recent most input sample at any given clock period is multiplied with all the filter coefficients. A set of intermediate results are generated in this case, and shared across all the multiplications in order to minimize the total number

of additions and subtractions using multiple constant multiplication techniques. Each such intermediate result in an MCM process corresponds to one of the common sub-expressions of the set of constants to be multiplied.

The common sub expression elimination for MCM reveals that the number of operators used to form the adder tree networks is very significant. The number of operators on an adder tree is determined by the number of input terms the coefficient uses from the network. For an N input adder tree, N-1 operators are required. Identify the resource minimization problem in the scheduling of adder tree operations for the FIR MCM block. The area and power consumption of the filters MCM blocks can be calculated and apply the MIP based algorithm for exact bit level resource optimization.

## 1.1 Related Work

The number of additions used to implement the coefficient multiplications. It determines the complexity of digital filters. Many approaches have been proposed in literature for reducing the number of adders in the multipliers of digital filters. Using coefficient partitioning method to implement low complexity digital filters with minimum number of full adders. While the optimization criterion in conventional low complexity filter implementation method is the number of adders, the focus of this method is to minimize the number of full adders required for each adder. The coefficient partitioning algorithm is combined with the pseudo floating point coefficient coding scheme and applied to optimize the common sub expression elimination methods. The full adder reduction achieved using this method is substantially higher for higher order filters.

The number of adders and critical paths in a multiplier block of a multiple constant multiplication based implementation of a finite impulse response filter can be minimized through common sub expression eliminate techniques. A two bit common sub expression can be located recursively in a non-canonic sign digit representation of the filter coefficients. To improve the elimination of a CS from the multiplier block of an FIR filter. It can be realized with fewer adders and logical depths as compared to the existing methods. This algorithm shows average logical operator with a comparative logic depth requirement.

The two reconfigurable FIR filter architectures, namely Constant shift method (CSE) and Programmable sub expression elimination (PSE). Among these approaches, techniques the best hardware reduction since it deals with the multiplication of one variable (input signal) with several constants (coefficients). The CSE techniques focus on eliminating redundant computations in multiplier blocks by employing the most common sub expressions consisting of two-nonzero bits. The PSM approach is based on the common sub expression elimination algorithm used. Unlike the CSM method where constant shifts are used, the PSM employs programmable shifters. The advantage of PSM over CSM is that the former architecture always ensures the minimum number of additions and thus minimum power consumption. The latency of the adder tree increases by using these approaches.

A new algorithm for digit serial FIR filter using CAD tool proposed in [2]. Little attention has been given to the digit-serial MCM design that offers alternative low complexity MCM operations at the cost of an increased delay. Designing digit-serial MCM operation with optimal area at the gate level by considering the implementation costs of digit-serial addition, subtraction, and shift operation. Since there are still instances with which the exact CSE algorithm cannot cope, the number of adders and logical operators still increases the latency and reduces the throughput of the multiple constant multiplications.

The resource minimization problem in the scheduling of adder tree operations for the FIR multiple constant multiplication block and used the mixed integer programming based algorithm for exact bit level resource

optimization. The power and area reduced in this approaches. After the problem identification of all these implementations to overcome that only we have to implement an efficient FIR filter and proposed the high speed carry select adder in the scheduling of adder tree MCM block for speech processing applications.

## II. PROPOSED METHOD

Digital Signal Processing (DSP) deals with the manipulation of digital signals using complex signal processing systems built from basic building blocks like filters. The proposed work implementation in the FIR filter in adder parts. This work evaluates the performance of the improved design in terms of delay and power. So we need to design high speed adders. The ripple carry adder part is replaced with the regular and improved carry select adder and the FIR Filter is implemented and the performance of the design is evaluated in terms of delay and power. The carry-select adder is a particular way to implement an adder, which is a logic element that computes the  $(n+1)$ -bit sum of two  $n$ -bit numbers. Which is the simple and fastest adder compared to all adders. The carry-select adder generally consists of two ripple carry adders and a multiplexer. Adding two  $n$ -bit numbers with a carry-select adder is done with two adders in order to perform the calculation twice, one time with the assumption of the carry being zero and the other assuming one. After the two results are calculated, the correct sum, as well as the correct carry, is then selected with the multiplexer once the correct carry is known. In the carry select adder initially when  $en=1$ , the output of the RCA(Ripple carry adder) is fed as input to the D Latch and the output of the D latch follows the input and given as an input to the multiplexer. When  $en=0$ , the last state of the D input is trapped and held in the Latch and therefore the output from the RCA is directly given as an input to the MUX without any delay.

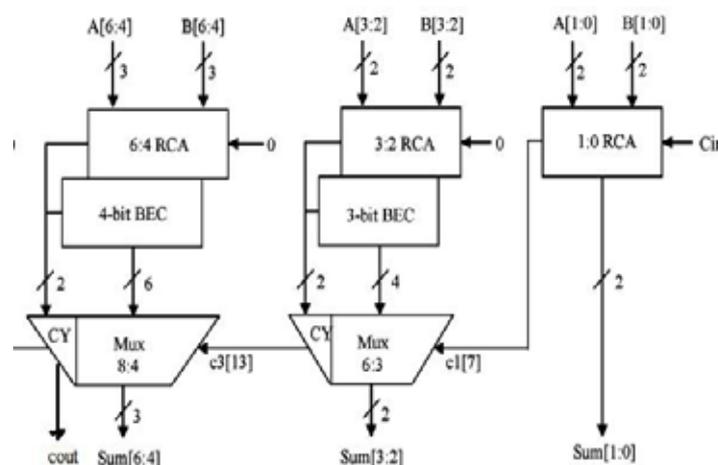


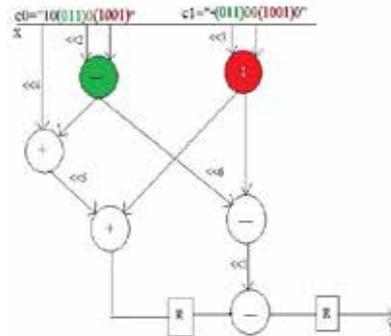
Fig 1. Carry select adder

Now the MUX selects the sum bit according to the input carry which is the selection bit and the inputs of the MUX are the outputs obtained when  $en=1$  and 0. The Adder tree formed using this carry select adder is explained in the below sections.

### 2.1 Adder Tree Scheduling

In the transposed FIR filter, the recent most input sample at any given clock period is multiplied with all the filter coefficients. A set of intermediate results are generated in this case, and shared across all the multiplications in order to minimize the total number of additions/subtractions using multiple constant multiplication (MCM) techniques. Each such intermediate result in an MCM process corresponds to one of the

common sub-expressions (CS) of the set of constants to be multiplied. The common practice of handling the summation of CS terms of each coefficient is to use the tree height minimization algorithm to produce a height optimum adder-tree. Tree-height minimization algorithm iteratively collapses the pair with smallest delays using an ADD/SUB to form a new term with delay, until a single term is reduced.



**Fig 2. Carry select adder tree scheduling**

Note that either a positive or negative sign is associated with each input term, which denotes whether the corresponding term should be added to or subtracted from the summation. These signs also determine whether an addition operation or a subtraction operation should be used when the algorithm collapses a pair of terms in the adder-tree based on the following rules.

- (1) If two input edges are of the same sign, an ADD will be used; otherwise, it will be a SUB.
- (2) The sign of the output edge is always the same as that of the “left” input edge.

Using these two rules, it is possible that the final term producing the summation result may carry a negative sign, such that a negation is needed after the adder-tree to correct the value. For an FIR filter, results from multiple adder-trees are accumulated by a structural adder register line. So the negation can be eliminated by replacing the structural adder with a sub tractor.

## 2.2 Cost Model

In order to quantify and minimize the hardware cost of the adder-tree, we model the cost of ADD/SUB operations in this section based on the carry select adder implementation, which is most area efficient and speedup process of scheduling the adder tree, it will be picked up by the hardware compiler whenever the timing allows. Without loss of generality, for a single ADD/SUB operation, the pair of its input operands may be of different bit-widths, and one of them is to be left shifted by certain bit positions. After the common sub expression terms are determined and the ADD/SUB network of non-redundant sub expressions (or terms) is formed, the product value corresponding to each of the coefficients is computed by an adder-tree that sums up its relevant terms. Two adder-trees are formed, for computing the product of a pair of coefficients using shifted versions of unique CS terms „1“, „10-1“ and „1001“ from the term-networks or sub expression-networks. We have developed the cost model of the ADD/SUB network by bit-level analysis, which could be reduced by suitable scheduling of operations on the adder-tree. We find that significant area, latency and power reduction. The cost calculation is done separately in three bit-segments. Starting from the least significant bit (LSB), the 1st segment covers the bit positions up to but not including the first bit of the shifted operand; the 3rd segment covers the bits corresponding to the sign extension bits of the sign extended operand; the 2<sup>nd</sup> segment takes the rest of the bit positions. The clock performance of the entire FIR filter is decided by the largest of the delays of all coefficients.

Assuming the delay of an ADD/SUB operator to be 1 unit, the delay of the constant multiplication by a coefficient can be simply measured by the number of ADD/SUB steps on a maximal path in the part of the network corresponding to the coefficient. The structural operator(s) of the adder-tree are not modeled on the binary tree. In linear phase FIR filters, which are used in most cases, a coefficient is accumulated twice at symmetric tap positions, thus corresponding to two structural operators. In a general FIR filter, an adder-tree corresponds to one structural operator. For a structural operator, the adder-tree output edge always Inputs to it as the right side operand, while the left input and output edges are on the accumulation line and their bit widths can be predetermined according to the upper/lower bound of accumulated values up to this coefficient tap. The left input edge is always "+" signed and carries 0 shift. The cost model for the logic depth and structural operator are calculated.

### 2.3 Mixed Integer Programming

We generally use Logic depth to describe the required ADD/SUB steps. For a coefficient whose logic depth is less than the filters logic depth, incrementing (relaxing) its logic depth may reduce the resource consumption. Given an algorithm which computes the adder-tree of the minimum resource on a given depth for a coefficient, if it is less than the filters logic depth, one can always try increasing by 1 and rescheduling onto a depth adder-tree for possible reduction of resource without degrading the filters clock performance. We describe the schedule of an adder-tree of an individual coefficient to minimize the hardware resource. As linearity is required in MIP, various techniques to transform modeling friendly non-linear expressions into linear equations and inequalities are indispensable.

## III. IMPLEMENTATION OF FIR FILTER USING CSA TREE

The performance of the multiple constant multiplication scheme is evaluated by implementing an FIR filter using carry select adder tree scheduling. The ripple carry adder replaced with the carry select adder to improve the performance of the filter. Which is the simple and fastest adder compared to all adders. The carry-select adder generally consists of two ripple carry adders and a multiplexer. The ripple carry adder part is replaced with the carry select adder and the FIR Filter is implemented and the performance of the design is evaluated in terms of delay and power.

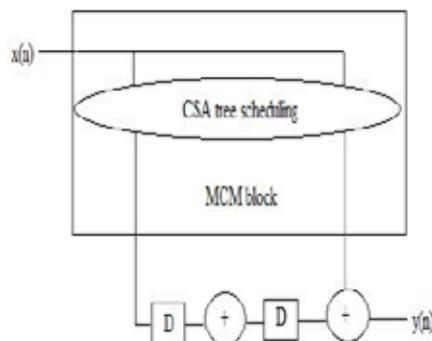


Fig. 3: Tranposed FIR filter

### 3.1 Single-Stage CSLA

The general expression to calculate the AOI gate counts of the n-bit proposed CSLA and the BEC-based CSLA of [6] and CBL-based CSLA of [7] and [8] are given in Table 1 of single-stage design. We have calculated the AOI

gate counts on the critical path of the proposed n-bit CSLA and CSLAs of [6]–[8] and used those AOI gate counts in (5b) to find an expression for delay of final-sum and output-carry in the unit of  $T_i$  (NOT-gate delay). The delay of the n-bit single-stage CSLA is shown in Table 1 for comparison.

**Table 1**

General Comparison of gate counts and delay of the proposed and existing CSLAS for single-stage design n: input bit-width

Design	AND-gate ( $N_a$ )	OR-gate ( $N_o$ )	NOT-gate ( $N_i$ )	final-sum ( $T_{fs}$ )	output-carry ( $T_{cout}$ )
Conventional	$14n - 4$	$7n - 3$	$5n - 1$	$\max(t, 3.5n + 2) + 4.5$	$\max(t, 3.5n + 1) + 4.5$
CSLA [6]	$11n - 2$	$5n - 1$	$7n$	$\max(t, 3.5n + 8.3) + 4.5$	$\max(t, 3.5n + 8.3) + 4.5$
CSLA [7]	$7n$	$4n$	$5n$	$4.5n + 1.8$	$4.5n + 1.8$
CSLA [8]	$14n - 7$	$8n - 4$	$9n - 4$	$9n + 1$	$9n + 1$
Proposed	$8n - 2$	$5n - 1$	$4n$	$\max(t, 3.5n + 2.7) + 8$	$\max(t, 3.5n + 2.7) + 3.5$

$t$  stands for delay of *input-carry*.  $t = 0$  for single stage adder design. Delay expressed in the unit of  $T_i$  (NOT-gate delay).

## VI. CONCLUSION

The adder tree scheduled in bit-level using CSA (carry select adder) are designed and are implemented in vhdl using Xilinx 13.2 ISE tool and the results are showed in terms of delay and power. The CSA proves to be the High Speed and Low Power multiple constant multiplication block. It is also implemented with FPGA. The MIP based algorithm used for bit level analysis. The performance of the MCM in terms of delay and power is evaluated by implementing an FIR Filter by using the CSA in the adder part and again it proves to be the High Speed and Low Power system. Thus the designed high speed FIR filter is used for speech signal processing applications.

## REFERENCES

- [1] Yu Pan and Pramod Kumar Meher, Senior Member, IEEE, “Bit-Level Optimization of Adder-Trees for Multiple Constant Multiplications for Efficient FIR Filter Implementation,” *IEEE transaction on circuits and systems-1: Regular papers*, vol. 61, no. 2, Feb 2014.
- [2] P. K. Meher and Y. Pan, “Mcm-based implementation of block fir filters for high-speed and low-power applications,” in *Proc. VLSI and System-on-Chip (VLSI-SoC), 2011 IEEE/IFIP 19th Int. Conf.*, Oct. 2011, pp. 118–121.
- [3] L. Aksoy, C. Lazzari, E. Costa, P. Flores, and J. Monteiro, “Design of digit-serial FIR filters: Algorithms, architectures, and a CAD tool,” *IEEE Trans. Very Large Scale Integration (VLSI) Syst.*, vol. 21, no. 3, pp. 498–511, Mar. 2013.
- [4] P. Prashanti, Dr. B. Rajendra Naik, “Design and Implementation of High Speed Carry Select Adder” *International Journal of Engineering Trends and Technology (IJETT) – Volume 4 Issue 9- Sep 2013*.
- [5] Paulchamy balaiah and Dr. Ila Vennila, “A Proficient Design of Hybrid Synchronous and Asynchronous Digital FIR Filter using FPGA”, *IJCSI International Journal of Computer Science Issues*, Vol. 9, Issue 4, No 1, July 2012, ISSN (Online): 1694-0814
- [6] M. B. Gately, M. B. Yeary, and C. Y. Tang, “Multiple real-constant multiplication with improved cost model and greedy and optimal searches,” in *Proc. IEEE ISCAS*, May 2012, pp. 588–591.
- [7] M. Kumm, P. Zipf, Faust, and C. H. Chang “Pipe line adder graph for MCM” *IEEE*, May 2012.

# PERFORMANCE COMPARISON OF GRADED INDEX OPTICAL FIBER USING DIFFERENT MODULATOR

Flintu K Kunjukunjan<sup>1</sup>, J.Geetha Ramani<sup>2</sup>, Chinchu Joseph<sup>3</sup>

<sup>1,3</sup>Department of Electronics and Communication Engineering, SNSCT, Coimbatore (India)

<sup>2</sup>Assistant Professor, Department of Electronics and Communication Engineering,  
SNSCT, Coimbatore (India)

## ABSTRACT

Optical fibers are widely used in fiber-optic communications, which permits transmission over longer distances and at higher bandwidths (data rates) than other forms of communication. The use and demand for optical fiber has grown tremendously and optical-fiber applications are numerous. The Graded Index optical fiber contains a core in which the refractive index diminishes gradually from the center axis out toward the cladding. The higher refractive index at the center makes the light rays moving down the axis advance more slowly than those near the cladding. In this paper  $Q$  factor of graded index fiber is calculated and the performance study is conducted with different modulator that is with MZ modulator and AM modulator. The result reveals the ability of graded index fiber is better by using MZ modulator.

**Keywords-**Multimode fiber, Graded index fiber, Intermodal dispersion,  $Q$  factor.

## I. INTRODUCTION

The per-channel light signals propagating in the fiber have been modulated at rates as high as 111 Gigabit per second(Gbit/s) by NTTT. The first type of polymer optical fiber(POF) is crofon was introduced in 1966.For high performance multimode polymer optical waveguide is used. Polymer optical fiber has same structure as silica optical fiber. For the fabrication of POF one type of optical polymer called polymethyl-methacrylate (PMMA) is used. The main characteristic of optical fiber is bandwidth and it is limited by signal dispersion. The signal dispersion determines the number of bits of information transmitted in given period of time. The intermodal dispersion is one of the dispersive properties of the fiber. In this light through optical fiber takes one or more path hence distance may vary from one path to another.

Fiber with large core diameter (greater than 10 micrometers) may be analyzed by geometrical optics. Such fiber is called *multi-mode fiber*, from the electromagnetic analysis . In a step-index multi-mode fiber, rays of light are guided along the fiber core by total internal reflection. Rays that meet the core-cladding boundary at a high angle (measured relative to a line normal to the boundary), greater than the critical angle for this boundary, are completely reflected. The critical angle (minimum angle for total internal reflection) is determined by the difference in index of refraction between the core and cladding materials. Rays that meet the boundary at a low angle are refracted from the core into the cladding, and do not convey light and hence information along the

fiber. The critical angle determines the acceptance angle of the fiber, often reported as a numerical aperture. A high numerical aperture allows light to propagate down the fiber in rays both close to the axis and at various angles, allowing efficient coupling of light into the fiber. However, this high numerical aperture increases the amount of dispersion as rays at different angles have different path lengths and therefore take different times to traverse the fiber. This paper undergoes Q factor calculation of graded index fiber when light pass through different modulation format. The system transmits data to 4km with power 10dbm. It was shown that the spectrum began to be shaped by eye diagram Analyzer, was used for the performance analysis. Optisystem13 software has been used for this simulation.

## II. BACKGROUND THEORY

Optical communication system using Plastic Optical Fibers(POF) have not reached their potential for a number of reasons, the rapid growth of glass optical fiber technology and because plastic optical fibers have low speed, short distance applications. Graded index plastic optical fiber is in great demand in customer premises to deliver high-speed services due to its high bandwidth, single-mode POF, optical amplification in plastic fibers, there are new POF materials with low loss and higher power and faster sources have been developed. The material most frequently used for the fabrication of POF is the thermoplastics PMMA (Polymethylmethacrylate). PMMA-SI-POF has a theoretical minimum attenuation of 106dB/km at 650nm.

The bond structure of PMMA is shown

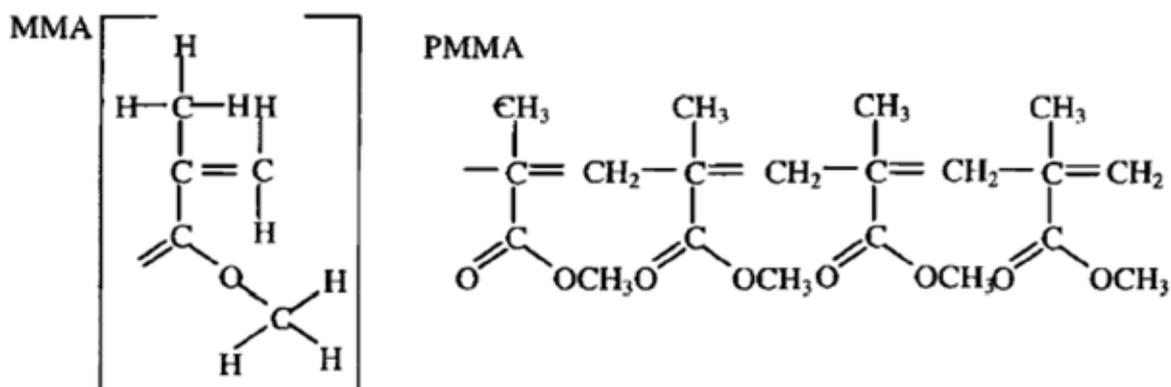


Fig.1 PMMA Bond Structure

### 2.1 Graded Index Fiber

The refractive Index profile describes the relation between the indices of the core and cladding. Two main relationships exist:

- (I) Step Index
- (II) Graded Index

The step index fiber has a core with uniform index throughout. The profile shows a sharp step at the junction of the core and cladding. In contrast, the graded index has a non-uniform core. The Index is highest at the center and gradually decreases until it matches with that of the cladding. There is no sharp break in indices between the core and the cladding.

By this classification there are three types of fibers :

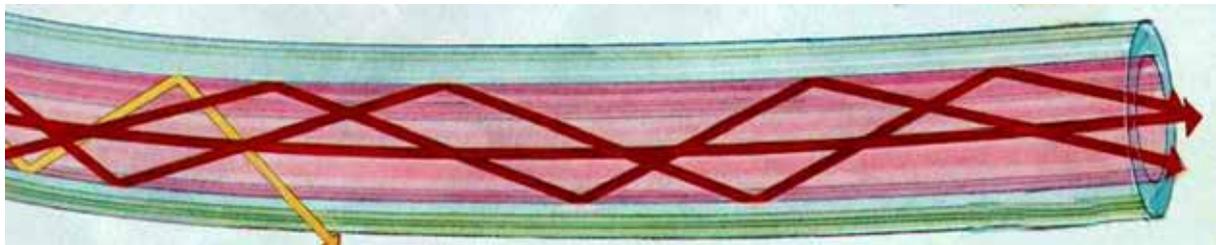
- (I) Multimode Step Index fiber (Step Index fiber)

(II) Multimode graded Index fiber (Graded Index fiber)

(III) Single- Mode Step Index fiber (Single Mode Fiber)

### 2.1.1 .Step-Index Multimode Fiber

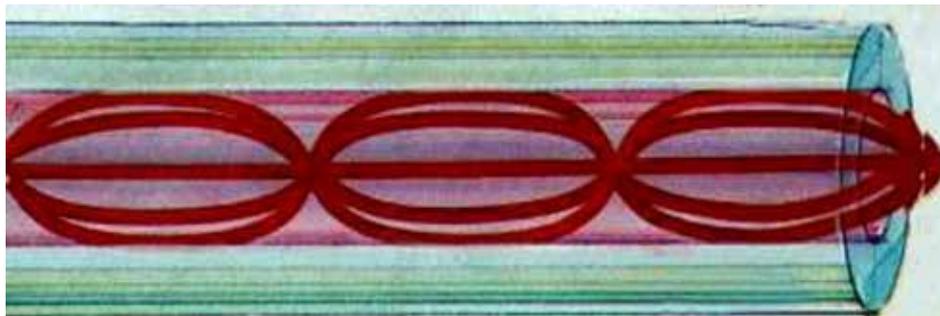
Step Index Fiber has a large core, up to 100 microns in diameter. As a result, some of the light rays that make up the digital pulse may travel a direct route, whereas others zigzag as they bounce off the cladding. These alternative pathways cause the different groupings of light rays, referred to as modes, to arrive separately at a receiving point. The pulse, an aggregate of different modes, begins to spread out, losing its well-defined shape. The need to leave spacing between pulses to prevent overlapping limits bandwidth that is, the amount of information that can be sent. Consequently, this type of fiber is best suited for transmission over short distances, in an endoscope, for instance



**Fig.2. Step Index Fiber**

### 2.1.2.Graded-Index Multimode Fiber

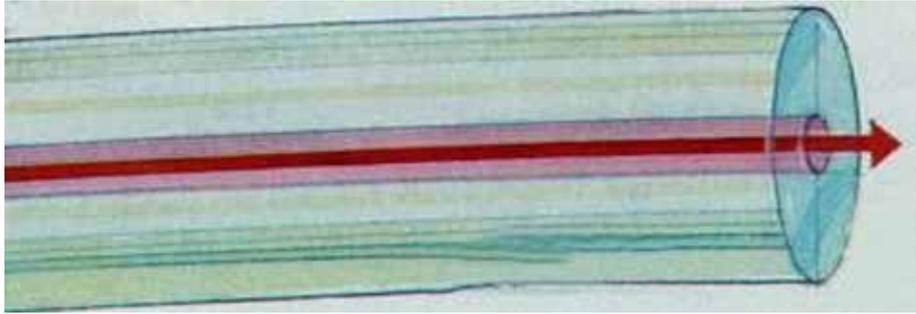
It contains a core in which the refractive index diminishes gradually from the center axis out toward the cladding. The higher refractive index at the center makes the light rays moving down the axis advance more slowly than those near the cladding



**Fig.3.Graded Index Fiber**

### 2.1.3. Single-Mode Fiber

It has a narrow core (eight microns or less), and the index of refraction between the core and the cladding changes less than it does for multimode fibers. Light thus travels parallel to the axis, creating little pulse dispersion. Telephone and cable television networks install millions of kilometers of this fiber every year.



**Fig.4.Single Mode Fiber**

## 2.2. Dispersion

Dispersion is the spreading of light pulse as it travels down the length of an optical fibre as shown in figure. The varying delay in arrival time between different components of a signal "smears out" the signal in time. This causes energy overlapping and limits information capacity of the fiber.

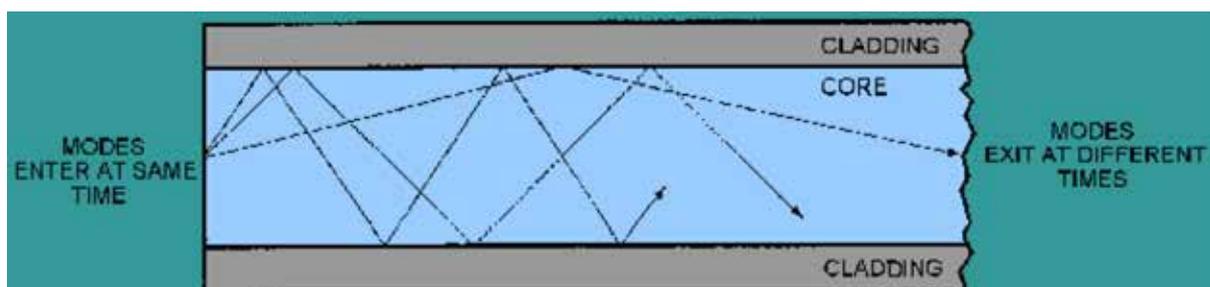
Dispersion limits the bandwidth or information carrying capacity of a fibre. The bit-rates must be low enough to ensure that pulses are farther apart and therefore the greater dispersion can be tolerated.

Dispersion of optical energy within an optical fiber falls into following categories:

- Ø **Intermodal Delay or Modal Delay**
- Ø **Intramodal Dispersion or Chromatic Dispersion**
- Ø **Material Dispersion**
- Ø **Waveguide Dispersion and Polarization –Mode Dispersion**

### 2.2.1. Intermodal Delay/ Modal Delay

Intermodal distortion modal delay appears only in multimode fibers. This signal distortion mechanism is a result of each mode having a different value of the group velocity at a single frequency. The amount of spreading that occurs in a fiber is a function of the number of modes propagated by the fiber and length of the fiber. Intermodal or modal dispersion causes the input light pulse to spread. The input light pulse is made up of a group of modes (MULTIMODE). As the modes propagate along the fiber, light energy distributed among the modes is delayed by different amounts. Modal dispersion occurs because each mode travels a different distance over the same time span. The modes of a light pulse that enter the fiber at one time exit the fiber different times. This condition causes the light pulse to spread. As the length of the fiber increases, modal dispersion increases.



### Fig.5. Intermodal Delay

#### 2.2.2. Intramodal Dispersion

Pulse spreading that occurs within a single mode Intra-modal dispersion occurs because different colors of light travel through different materials and different waveguide structures at different speeds Also called GROUP VELOCITY DISPERSION (GVD ). It can Occurs in all types of fibers Two main causes : *Material dispersion*  
*Waveguide dispersion*

##### A. Material Dispersion

Arises from variations of the refractive index of the core material as a function of wavelength Different wavelengths travel at different speeds in the fiber material and hence exit the fiber at different times

##### B. Waveguide Dispersion

Arises because a Single Mode Fiber confines only 80% of the optical power to the core The other 20% tends to travel through the cladding and hence travels faster This results in spreading of the light pulses .The amount of dispersion depends on the fiber design and the size of the fiber core relative to the wavelength of operation .In multimode fibers, waveguide dispersion and material dispersion are basically separate properties. Multimode waveguide dispersion is generally small compared to material dispersion and is usually neglected

### III.SIMULATION SETUP

The optisystem software is very important will be used to model and simulate fiber optic system. The transmitted of optical system consist of Pseudo –Random Bit used to generate sequence random of bits (0 or 1), NRZ pulse generator has an advantage on controlling bandwidth. This is due to the characteristic of the generator that the returning signals to zero between bits. Pseudo-random bit sequence generator is used to scramble data signal in terms of bit rates[16], Mach Zender- Modulator has two inputs and one output. Then the input signal is modulated with semiconductor laser that is represented by spatial Continuous Wave (CW) laser through Mach-Zehnder modulator. The spatial Continues laser diode (CW) to generate optical signals supplies input signal with 850 nm wavelength and input power variable from 0- 30dBm which is externally modulated. The pulses are launched into the fork 1xN which distributed power into three branch

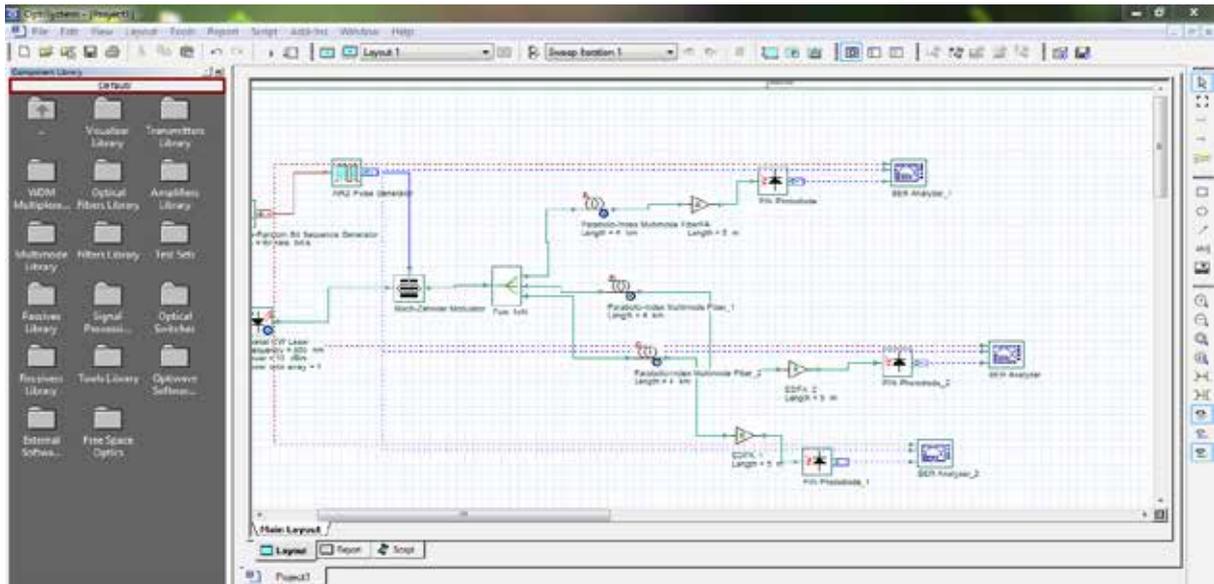


Fig.6. Simulation Set Up

In this circuit diagram instead of MZ modulator an AM modulator is inserted and analyze the performance.

IV.RESULTS

The eye diagrams and results of maximum Q. factor is shown in the figure.

Table 1

Modulator	Q Factor
MZ Modulator	39.0505
AM Modulator	10.8114

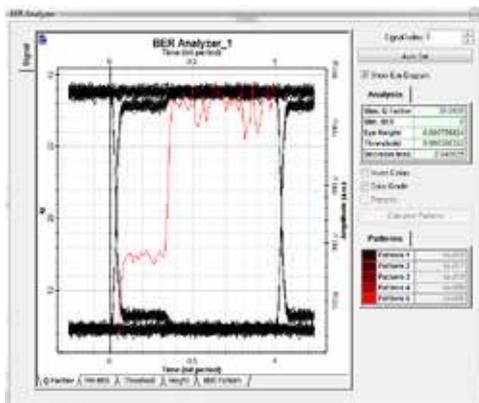


Fig.7. EYE Diagram for MZ Modulator

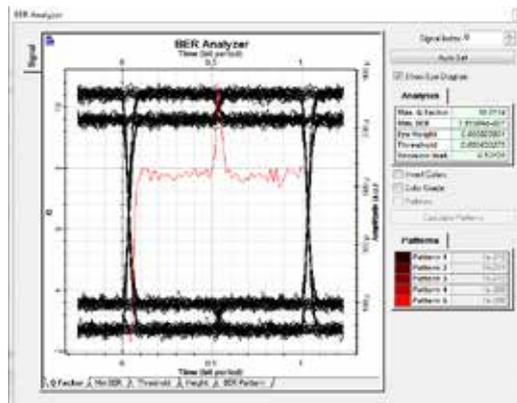


Fig.8. EYE Diagram for AM Modulator

## V. CONCLUSION

In this paper, the design study to improve Q factor of transmission system. The results are valuable for improving system performance by using graded index multimode fiber. The Q factor is compared with different modulator. From the simulation results, concluded that Q factor is better for MZ modulator than AM modulator. For MZ modulator the Q factor is 39.0505 and Q factor for AM modulator is 10.8114.

## REFERENCES

- [1] Mahmoud M. Al-Quzwini “ Design and Implementation of a Fiber to the Home FTTH Access Network based on GPON” International Journal of Computer Applications (0975 – 8887) Volume 92 – No.6, April 2014
- [2] Dheyaa Jasim Kadhim, Nahla Abdul-Rahman Hussain “ Design and Implementation of a Practical FTTH Network” International Journal of Computer Applications (0975 – 8887) Volume 72– No.12, June 2013
- [3] Brian Stack, Matthew Sargent, Tom Dooner, and Mark Allman “A First Look at 1 Gbps Fiber-To-The-Home Traffic” TR-12-009 August 2012. International computer science institute
- [4] Arvind Kumar, Jaiswal ,SandeepSingh ,Mukesh Kumar “Transmission of Voice, Video and Data in Fiber-To-The-Home (FTTH) Networks Using OFDM” Volume 2, Issue 5, May 2012 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering
- [5] Prof Jai Karan Singh, HOD EC, SSSIST Sehore, Prof. Mukesh Tiwari, Dean Academic, SSSIST Sehore, Prof Sanjay Rathor, SSSISTSehore, Mr. Vikas Shrivastav, SSSIST Sehore (D.C) “Simulation of 2gb/s Downstream Transmission Performance of Gpon” International Journal of Advanced Technology & Engineering Research (IJATER)-2012
- [6] Aswir Premadi, Boon Chuan, Kasmiran Jumari ,Mohammad Syuhaimi Ab-Rahman, and]“Physical Layer Monitoring in 8-branched PON-based i-FTTH” @2010 IEEE
- [7] Lluís Gutierrez<sup>1</sup>, Marilet de Andrade<sup>2</sup> and Sebastià Sallent<sup>3</sup> Dpt. of Telematics, “New trends on Optical Access Networks: DBAs for 10G EPON and Long-Reach PON” Polytechnic University of Catalonia (Spain) E-mail: (lluis.gutierrez, marilet)

# A REVIEW ON DIFFERENT METHODS FOR SECURITY IN E-COMMERCE TRANSACTIONS

Meenu Alex<sup>1</sup>, Anu Kunjumon<sup>2</sup>, Smitha Jacob<sup>3</sup>

<sup>1,2</sup> Department of Computer Science , St. Joseph's College of Engineering & Technology ,  
Palai, Kerala (India)

<sup>3</sup> Assistant Professor, Department of Computer Science ,  
St. Joseph's College of Engineering & Technology, Kerala (India)

## ABSTRACT

*With the rapid growth of Internet technology, people are rely upon network for all kinds of transactions. As a result E-commerce is widely developed. Security is the major concern of these transactions. For secure transactions, proper security and privacy protection should be provided. To authenticate remote users, we need a more secure remote password authentication scheme to resist different kinds of attacks and vulnerabilities. So far, different kinds of authentication mechanisms are developed to mitigate the risk of network attacks. Passwords are the most commonly used authentication mechanism. But passwords can be easily compromised and are susceptible to different attacks. So some advancement in the password mechanisms. These include smart card with check digits, dynamic ID based authentication etc. The problems with these methods are guessing attack, impersonation attack, smart card theft attack etc. Different password schemes can be merged together to form a single and more secured password scheme. Online biometric solution is the most secure authentication mechanism that has ever identified. Because biometric information can't be stolen or compromised. This seminar proposes two methods to enhance security in E-commerce transactions. For least sensitive transactions SSL protocol together with a fingerprint sample can be used. For high sensitive transaction 3-D secure protocol together with iris sample with the help of one time password is used. The iris is the most particular feature of a human being.*

**Keywords :** *Authentication, Biometric Authentication, E-Commerce, Matching , Template,*

## I. INTRODUCTION

Two problems exist in applying biometric characteristics for remote authentication. Firstly, some biometric characteristics may be easily obtained and they can never be changed, which makes these biometric characteristics unreliable as encryption keys. Secondly, the biometric capture devices are located remotely. The server is not possible to check whether the device is capable of verifying that a person is alive or not. So the best approach is to combine biometrics with passwords and smart cards to construct a secure three-factor authentication scheme. In such a system, only a user with the correct password, biometric characteristics, and data in his smart card is allowed to login to the remote server. Each of the three factors passwords, smart cards, and biometrics, has its own merits and demerits that complement one another.

A three-factor authentication scheme uses passwords, smart cards and biometrics to provide high-security authentication. Most existing schemes depend on smart cards in order to verify biometric characteristics. The biometric data of the user is not shared with the server. But the demerit is that the remote server should trust the smart card to perform proper authentication which in turn leads to various attacks. Different researchers studied various types of passwords, their benefits and drawbacks, how they are vulnerable to different types of attacks.

## II.PASSWORD AUTHENTICATION WITH INSECURE COMMUNICATION

In 1981, L. Lamport done a work on Password Authentication [1] with insecure Communication. He found that in remotely accessed computer systems, a user identifies himself to the system by sending a secret password. There are three ways an intruder could learn the user's secret password and then impersonate him when interacting with the system:

1. By obtaining access to the information stored inside the system like reading the system's password file.
2. By seizing the user's communication with the system like eavesdropping on the line connecting the user's terminal with the system.
3. By the user's disclosure of his password like choosing an easily guessed password.

The third possibility cannot be prevented by any password protocol, since two individuals presenting the same password information can be distinguished by the system. The first weakness can be eliminated by a one way function to encode the password. A one way function is a mapping  $F$  from some set of words into itself such that :

1. Given a word  $x$ , it is easy to compute  $F(x)$
2. Given a word  $y$ , it is not feasible to compute a word  $x$  such that  $y = F(x)$ .

To prevent the second weakness one must use a sequence of passwords  $x_1, x_2, \dots, x_{1000}$  where  $x_i$  is the password by which the user identifies himself for the  $i^{\text{th}}$  time. Since users are using commercial services through networks, the user authentication protocol has been regarded as a most important security issue. However, many vulnerabilities are present in the user authentication protocol due to the careless password management and not matured attack techniques. Recently, with the appearance of smart card, these vulnerabilities were reduced.

## III.CRYPTANALYSIS OF A NEW DYNAMIC ID-BASED USER AUTHENTICATION SCHEME TO RESIST SMART-CARD-THEFT ATTACK

In 2004, Das et al. [2] proposed ID-based authentication protocol to mitigate vulnerabilities of the password-based authentication protocol. This authentication protocol has some advantage that it allows a user to change a password easily. By using a hash function, it requires low computation cost. Also, the server does not need to maintain the password verification table. Wang et al. explained that ID based authentication protocol allows an attacker to complete the authentication without knowing the password. To improve the earlier protocol, Wang et al. proposed the secure protocol using dynamic ID and the mutual authentication between user and the remote

server. The Wang et al.'s protocol is not secure against the message alteration attacks and impersonation attacks. In Wang *et al.*'s protocol, an attacker can be authenticated as a valid user without knowing the ID and password. If an attacker eavesdrops the log-in requesting message  $\langle ID_i, CID_i, N_i, TS \rangle$  the valid user  $U_i$  transmitted to remote server  $S$ . The smart card is equipped with tamper resistance, many researches have shown that the secret values stored in smart card can be extracted by executing the differential power attacks and fault attacks. Malicious attacker  $U_m$  participating in protocol can execute the impersonation attack because  $U_m$  can extract the secret value of  $\gamma$  server stored in his or her smart card by executing the differential power attacks and fault attacks.

#### **IV.A ROBUST AND FLEXIBLE BIOMETRICS REMOTE USER AUTHENTICATION SCHEME**

In May 2012, Eun-Jun Yoon and Kee-Young Yoo [4] presented a robust and flexible biometrics remote user authentication scheme. The proposed scheme has several important features and advantages as follows :

- (1) It uses a small communication round to optimize the computation cost of each participant.
- (2) By using bit-wise XOR operation and collision-free one-way hash functions, it achieves cryptographic goals.
- (3) It not only is secure against cryptographical attacks such as replay attack, guessing attack, parallel session attack, reflection attack, insider attack and impersonation attack. It also provides mutual authentication and secure password change function without helping of the remote server. ‘

This scheme is very useful in smart card-based Internet and wire/wireless communication environments to access remote information systems. To prevent the proposed parallel, it uses the input values of  $C1$  and  $C2$  in different formats when performing the login and authentication phases. To prevent the proposed privileged insider attack, this scheme uses a random nonce  $n$  to protect the password  $pw_i$  in the registration phase. There are four phases in this scheme including registration, login, authentication, and password change like Khan-Zhang's scheme.

#### **V.REMOTE USER AUTHENTICATION SCHEME USING SMART CARDS WITH CHECK DIGITS**

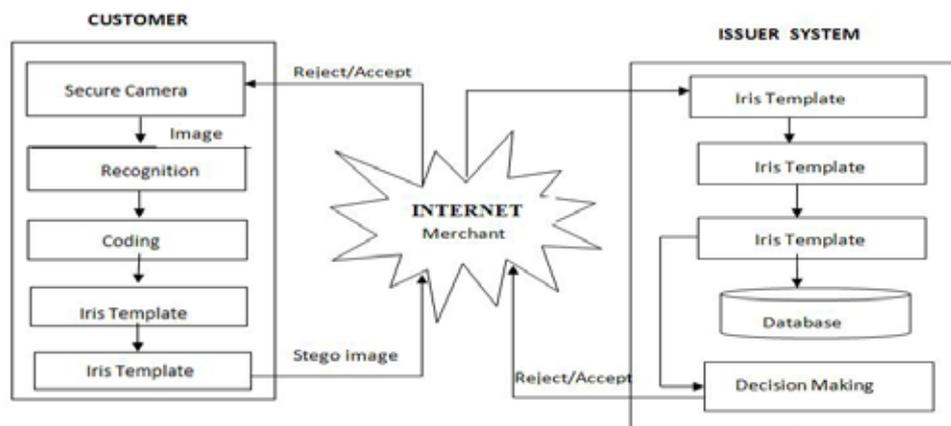
In 2000, Hwang et al. [9] proposed a new remote user authentication scheme using smart cards. Chan and Chang showed that the masquerade attack is successful on this scheme. Further in 2003, Leung et al [11] showed that this modified scheme is still susceptible to the attack proposed by Chan and Cheng. They clearly showed that the extended attack proposed by Chang and Hwang [12] also works well. The new remote user authentication scheme is composed of the initial phase, login phase, registration phase, and the authentication phase. This scheme is the modified form of the Hwnag-Li's scheme. For each registered identity, it uses one more function CK to generate the check digit. Here, only the AS can generate a check digit and a valid identity.

## VI. EFFICIENT IRIS AUTHENTICATION USING CHAOS THEORY-BASED CRYPTOGRAPHY FOR E-COMMERCE TRANSACTIONS

Arian Rahimi *et al* proposed an idea about an Efficient Iris Authentication Using Chaos Theory-based Cryptography for E-commerce Transactions[6]. Credit cards are the prominent method for purchasing goods and services through online. The information stored in the credit card is transmitted over the internet. So it does not provide the same level of security as phone, fax or mail. Security threats are the most important risk in online shopping. Therefore proper mechanisms to ensure secure online payment are significant to the success of purchasing over the Internet.

Credit card theft on the Internet is a most common problem. Any person who is having access to a credit card can purchase anything over the Internet if he knows the credit card number and expiration date. Credit cards were designed to depend on physical signatures for authentication which is a mechanism that is performed useless in e-commerce. The online buyer does not need to provide a physical card which may contain additional security features.

Biometric technique is used for preventing fraudulent appropriation and false authentication. Iris is recommended as the most unique feature. Iris recognition is now considered as the best and most exact solutions to security problems for human identification. This architecture involves two schemes: Image processing and secure template transmission scheme. Biometric template transmission scheme and Iris recognition schemes are used to decrease online credit card frauds.



**Figure 1.1 :Architecture of online credit card transaction**

While giving out a credit card, the Iris details of an individual is stored in the issuing agency's database along with the credit card number and other personal details. A software is presented in all the client systems. While doing e-commerce transactions, the Iris image of the individual was captured then it is encrypted and sent along with the name, credit card number and expiration date. The client system is equipped with a web camera. At the time of transaction the Iris image of the customer is obtained using that specific web camera. The captured Iris image first undergoes preprocessing to remove the distortions. Then it is normalized and enhanced. Finally, the key features of the Iris are extracted using a high performance algorithm.

Every Iris recognition algorithm is composed of 3 main sections:

1. The image is preprocessed in order to detect distortions and separate Iris from the whole image
2. Iris patterns are clipped as a code
3. Decision is obtained by performing matching.

There are two basic problems in encryption:

1. Hackers are capable to find ways to crack encryption.
2. A hacker is free to make a new copy of the data by decrypting the legitimate copy of the data.

RSA is used as a cryptographic method to encrypt biometric templates. RSA is a well known, established encryption algorithm but it confronts with some attacks like adaptive chosen cipher text, timing attacks, and branch prediction analysis attack. An attack against RSA can occur when the message is short. Since the volume of the "Iris Database" is too large and the biometrics data of every person is saved into it, this threat exists in iris template encryption.

## **VII. SECURE THREE-FACTOR AUTHENTICATION SCHEME WITH PRIVACY PROTECTION ON BIOMETRICS**

Biometric data can be used as a key to encrypt a secret, and the idea preserves the privacy of biometric templates [5]. A biometric template with an embedded secret, called a sketch can be public. If a sketch is secure, someone who obtains or keeps the sketch cannot recover the original biometric template or the secret without a successful biometrics matching. The algorithm of extracting the secret from the sketch is to match the biometrics. It is required to input the sketch and a biometric sample. If the biometric sample matches the template in the sketch, the secret will be correctly extracted. Some approaches, such as error-correcting codes, fuzzy commitments, fuzzy vaults, and fuzzy extractors, which support fuzzy-based key extraction and tolerate the jitters of biometric data, have been proposed in the literature, where the approaches of and can take binary biometric data, like irises, as keys for encryption/decryption.

In biometric-based authentication, the above idea should be performed on the client side rather than on the server side for the protection of biometrics against the server. If we want the server to perform the secret-extracting algorithm for truly three-factor authentication, it should know the input biometric sample. And after executing the secret-extracting algorithm, the server also obtains the embedded secret. The secret can recover the original biometric template with the public sketch. Therefore, the technique is not sufficient to solve our problem because it does not consider privacy protection for biometric data against the server when the operations about biometrics are performed in the server's side.

This scheme matches biometrics in the server and considers the strong privacy of biometrics. Since all of the above previous schemes perform the checking of biometrics in the cards, the server must trust the smart cards to properly handle the task. An adversary can skip the checking process in the smart cards without the remote server noticing. To address this problem, we verify all of the users' biometrics in the server. Such an approach must consider the fact that users are not willing to reveal their biometrics to others, even to the servers, because their biometric characteristics can never be changed and may also be applied to other activities. Therefore, this scheme also protects the user's biometric data from the server. In addition to the security and privacy issues, the

proposed scheme is tailored to achieve the low-computation requirements of smart card.

## VIII.CONCLUSION

The biometric-based authentications are widely used now due to its uniqueness and non-repudiation. Just like any other mechanisms, the systems developed by each researcher is not fully developed and fully error-proof. Even then, these technological babies are hatching themselves in the research incubators till date. The day is not far when we find one among these systems or a derivative system taking over and giving cent percent protection to all our digital data. This system is quite unfamiliar to the common users now; the study proves them strong and efficient to be brought into the scene of computational real soon.

## REFERENCES

- [1] L. Lamport,1981; Password authentication with insecure communication, Communications of the ACM 24 (11) 770–772.
- [2] Fengtong Wen, Dianli Guoand Xuelei Li; Cryptanalysis of a New Dynamic ID-based User Authentication Scheme to Resist Smart Card-Theft Attack, Appl. Math. Inf. Sci. 8, No. 4, 1855-1858 (2014)
- [3] A.K. Awashti 2004; A dynamic ID-based remote user authentication scheme, Transactions on Cryptology I (2) 15–16,2004
- [4] Eun-Jun Yoon and Kee-Young Yoo; A robust and flexible biometrics remote user Authentication scheme, International Journal of Innovative Computing, Information and Control Volume 8, Number 5(A), May 2012
- [5] Chun-I Fan and Yi-Hui Lin; Provably Secure Remote Truly Three-Factor Authentication Scheme With Privacy Protection on Biometrics, IEEE Transactions on Information Forensics and Security, vol. 4, no. 4, December 2009
- [6] Arian Rahimi, Sharhriar Mohammadi and Rozita Rahimi; An Efficient Iris Authentication Using Chaos Theory-based Cryptography for E-commerce Transactions
- [7] S. Mohammadi, S. Z. Hosseini, 2010;Virtual password using Runge- Kutta method for internet banking, Second International Conference on Communication Software and Networks
- [8] C.C. Chang, K.F. Hwang, 2003; Some forgery attacks on a remote user authentication scheme using smart cards, Informatics 14 (3) 289–294.
- [9] M. S. Hwang and L. H. Li; A new remote user authentication scheme using smart cards, *IEEE Trans. Consumer Electron.*, vol. 46, No. 1, pp. 28-30, 2000.
- [10] C. C. Chang and T. C. Wu, Remote password authentication with smart cards, *IEEE Proceedings*, vol. 138, no. 3, pp. 165-168, 1993
- [11] K.C Leung, L. M. Cheng, A. S. Fong, and C. K. Chan, Cryptanalysis of a modified remote user authentication scheme using smart cards,IEEE Trans. Consumer Electron., vol. 49, No. 4, pp. 1243-1245, Nov 2003
- [12] C. C. Chang and K. F. Hwang, Some forgery attack on a remote user authentication scheme using smart cards, *Infomatics*, vol. 14, no. 3, pp.89 - 294, 2003

- [13] J. K. Lee, S. R. Ryu, and K. Y. Yoo, "Fingerprint-based remote user authentication scheme using smart cards," *Electron. Lett.*, vol. 38, no. 12, pp.554–555,2002.
- [14] C. H. Lin and Y. Y. Lai, "A flexible biometrics remote user authentication scheme," *Comput. Standard Interfaces*, vol. 27, no. 1, pp. 19–23,2004.

# A STUDY OF FACTORS THAT ARE AFFECTING ONLINE SHOPPING BEHAVIOR OF CONSUMERS IN INDIA

<sup>1</sup>Mr. Debansu Chatterjee, <sup>2</sup>Asst. Prof. Indrajit Ghosal

<sup>1</sup>Research Scholar

Applied PhD in WBUT awaiting for enrollment in Management.

Faculty of EDI (EDI-Entrepreneurship Development Institute, Govt. of India,( Kolkata)

<sup>2</sup>Asst. Professor , Department in IT,

Institute of Management Study (IMS), Kolkata

Applied PhD in WBUT awaiting for enrollment in Management.

## ABSTRACT

*On-line shopping is a recent phenomenon in the field of E-Business and is definitely going to be the future of shopping in the world. Most of the companies are running their on-line portals to sell their products/services on-line. Though online shopping is very common outside India, its growth in Indian Market, which is a large and strategic consumer market, is still not in line with the global market. The potential growth of on-line shopping has triggered the idea of conducting a study on on-line shopping in India. The present research paper has used Qualitative and Quantitative research methods to study the impact of Demographic factors of consumers on on-line shopping parameters like satisfaction with on-line shopping, future purchase intention, frequency of on-line shopping, numbers of items purchased, and overall spend on on-line shopping. The data was collected through Questionnaires. The results of study reveal that on-line shopping in India is significantly affected by various Demographic factors like age, gender, marital status, family size and income. The results of the study could be further used by the researchers and practitioners for conducting future studies in the similar area.*

**Keywords :** *Demographic Factor, Age, Gender, Occupation, Income, On-Line Shopping.*

## I. INTRODUCTION

Currently the retail industry in India is accelerating. It is not widely accepted as the way it is accepted in Asian counterparts. India is excited to grow to be a most important player in the retail market. Since India is a developing nation, it is still not prepared for it. And they are leaving no stone unturned to become the best in retail Industry. Also India with a high on cross culture factor, it allows different companies bringing in variety of products Targeting different consumer segments.

According to the Global Retail Development Index 2012, India ranks fifth among the top 30 emerging markets for retail. There are many factors contributing to the boom in this sector. Some of them are increased consumerism with the ability to afford luxury items. Even there is increase in the spending power of Indians. Indian's are travelling abroad and they are exposed to different cultures, their way of living, their style and

thereby more brands. Worldwide retailers eye the Indian market because in our culture there is lots of give And take for marriages, festivals and other important events. Hence it's imperative for people to spend on gifts as a part and parcel of life irrespective of caste and creed and economic status. As there is no limitation to the spending power, Cities like Chandigarh, Coimbatore, Pune, Kolkata, Ahmadabad, Nagpur, Cochin, Hyderabad etc. has fine market for about all product or service which offers a good proposal for a brand to penetrate Indian market

## **II. ONLINE RETAIL SCENARIO IN INDIA**

The fact that world is fast dwindling into a universal community for the reason that of Internet and other communication mediums is not completely reflective in the Indian context. While developed and fast developing countries have understood the influence of Internet, India is still ignorant and trying to deal with up with the technological developments. The retail industry campaign has been repainted by Internet and the rules of the game in retailing are fast altering. The western culture has affected each and every aspect of our Indian Society. The life is becoming fast not only in metros but also in the normal cities. The number of nuclear families is increasing and both husband and wife are working, as they have less time to go to the market for purchasing every now and then.

Some other reasons like these, say shortage of time, traffic jams, late working hours, versatility of plastic money and above all the approach of internet at the door step of whosoever desires it. Online retailers have improved their service and consumers have found it convenient. There is been transformation in payment mode as well. From advance payment it is moved to cash on delivery (COD). Even in case of delivery pattern is changed. From fixed delivery timings it is moved on to convenient delivery timings at the choice of the customer. India has opened the doors for foreign direct investment in retail.

Indian market is dominated by unorganized players but there is potential in the area of retail players as well. Entry of bigger players such as Big Bazaar, More etc. To even in the rural areas is paved the way of growth in its sector. There is been rise in the increase of Indian middle class people due to rapid economic growth. Though the population of using Internet in India is low in terms of by and large percentage of total population but, in total numbers it is very high. This gives enormous opportunities for various ecommerce sites to get in touch with this segment. The population accessing in India is the age group from 1845. Irrespective of this fact online retailing forms a meager 0.08% of the entire Indian retail markets. Market players have to be more positive, revolutionizing and pioneering in their approach and offering to make serious inroads. In the current market, bulk of online sales is in a range of items. This market in India needs to leap to the next level.

## **III. LITERATURE REVIEW**

**Solomon, 1998** in his study "Consumer behavior is the study of the processes involved when an individual selects, purchases, uses or disposes of products, services, ideas, or experiences to satisfy needs and desires". In view for the Internet to spread out as a retail channel, it is imperative to realize the consumer's mind-set, intention and conduct in light of the online buying practice: i.e., why they employ or falter to use it for purchasing? Consumer attitudes seem to have a significant influence on this decision.

**(Schiffman, Scherman, & Long, 2003 in his study researched that** "yet individual attitudes do not, by themselves, influence one's intention and/or behavior. Instead that intention or behavior is a result of a variety of attitudes that the consumer has about a variety of issues relevant to the situation at hand, in this case online

buying. Over time the Internet buyer, once considered the innovator or early adopter, has changed. While once young, professional males with higher educational levels, incomes, tolerance for risk, social status and a lower dependence on the mass media or the need to patronize established retail channels (**Ernst & Young, 2001; Mahajan, Muller & Bass, 1990**), **Sultan and Henrichs (2000)** in his study concluded that the consumer's willingness to and preference for adopting the Internet as his or her shopping medium was also positively related to income, household size, and innovativeness. **Vijay, Sai. T. & Balaji, M. S. (May 2009)**, revealed that Consumers, all over the world, are increasingly shifting from the crowded stores to the one-click online shopping format. However, in spite of the convenience offered, online shopping is far from being the most preferred form of shopping in India. A survey among 150 internet users, including both users and non-users of online shopping, was carried out to understand why some purchase online while others do not. The results suggested that convenience and saving of time drive Indian consumers to shop online; while security and privacy concerns dissuade them from doing so. **The work of Kim and Park (2005)** using U.S. samples suggests that their positive attitudes as well as willingness to search for pre-purchase information leads to a strong likelihood that they will buy online. Online shoppers are required to have computer skills in order to use the Internet for shopping. Hence, those who are not comfortable with using the computer, will likely do their shopping at the traditional store, modern shop, or discount store (**Monsuwe , 2004**) because it will be faster shopping there than in the Internet shop. **Goldsmith and Flynn (2004)** state that the home catalog is another traditional selling channel where people can shop at home because of the varieties of products offered in the catalog. They can order through the phone or by mail. It is convenient except that they are not able to touch and feel products before purchasing. (**Bhattacharjee Sarathi Partha.,et al., 2012;**) **Zhang Lingying, Tan Wojie, Xu yingcong,Tan Genlue (2012)** searched the consumers' perceived risk exists in every phase of the overall process of B2C (Figure-1) electronic commerce while buying online, so it should be considered, because it influences e-consumers buying behavior. (**Kumari Renu.,2013;**) Use of technology has opened new doors and opportunities that enable for a more convenient lifestyle today. Variety of products, quicker services and reduced price are the three significant ways in which online shopping influenced people in India and world as a whole. **Patna (2013)**, investigates the relationship between globalization, ecommerce adoption or acceptance that lead to business performance and effectiveness. Through privacy and security policies, developers are doing their best to put an end to this unethical practice. That will pay the way for its success.

#### IV. RESEARCH OBJECTIVES

To identify the impact of Demographic factors \* impacting on-line shopping behavior\*\* of consumers with special emphasis on

\* Age, Gender, Education, Income.

\*\* Possession of internet, Frequency of online purchase, Motivation drives for online purchase.

#### V. RESEARCH HYPOTHESES

H0: Demographics Factors of Consumers not creating any Impact on the Online Shopping Behavior of the Consumers.

H1: Demographics Factors Of Consumers Significantly Impact The Online Shopping Behavior Of The Consumers.

H1.1: Age of consumer significantly impacts online shopping parameters\* of consumer.

H1.2: Gender of consumer significantly impacts online shopping parameters\* of consumer.

H1.3: Marital Status of consumer significantly impacts online shopping parameters\*of consumer.

H1.4: Income of consumer significantly impacts online shopping parameters \*of consumer.

\* Possession of internet, Frequency of online purchase, Motivation drives for online purchase.

## VI. RESEARCH METHODOLOGY

### Methods of data collection

A combination of Interview method and Questionnaire method is used to collect data from the respondents.

### Sample Design

Under sample design the method of Random Sampling is been used to collect data from the respondents.

### Sample Size

Sizes of 70 respondents are taken for the collection of the data.

### Data Analysis and Interpretation

In data analysis and interpretation method of “ANOVA” is used to analyze the data by using the minitab 16 Software and SPSS.

## VII. HYPOTHESES TESTING

H1.1: Age of consumer significantly impacts on possession of internet connection of consumers.

## VIII. CONSUMERS’ RESPONSE VARIATIONS FOR ONLINE BEHAVIOR \* ACROSS DIFFERENT AGE GROUPS

(\* Possession of internet, Frequency of online purchase, Motivation drives for online purchase)

Dependent variables	Mean	Std. Deviation	F-Value	Sig.
Possession of internet	1.11	.320	7.125	.000
Frequency of online purchase	3.41	.577	4.561	.006
Motivation drives for online purchase	2.23	1.024	1.493	.225

The results show that Age as one of the variables of demographics factor does impact Possession of internet and Frequency of online purchase online shopping behavior of consumers.

## IX. CONSUMERS’ RESPONSE VARIATIONS FOR ONLINE BEHAVIOR \* ACROSS DIFFERENT Gender GROUPS

(\* Possession of internet, Frequency of online purchase, Motivation drives for online purchase)

Dependent variables	Mean	Std. Deviation	F-Value	Sig.
Possession of internet	1.11	.320	7.125	.000
Frequency of online purchase	3.41	.577	4.561	.006
Motivation drives for online purchase	2.23	1.024	1.493	.225

The results show that **Gender** as one of the variables of demographics factor **does impact Possession of internet and Frequency of online purchase** of consumers. The result may be documented to the reason that females are more impulsive buyers as compared to males and are more likely to be attracted towards the promotional schemes offered by the online retailers and therefore gender has a significant impact on frequently of on-line shopping. Rising Working woman concept has also enhanced it. Females also believe in status quo and emulate others in shopping.

#### X. CONSUMERS' RESPONSE VARIATIONS FOR ONLINE BEHAVIOR \* ACROSS DIFFERENT INCOME GROUPS

(\* Possession of internet, Frequency of online purchase, Motivation drives for online purchase)

Dependent variables	Mean	Std. Deviation	F-Value	Sig.
Possession of internet	1.11	.320	2.927	.027
Frequency of online purchase	3.41	.577	5.397	.001
Motivation drives for online purchase	2.23	1.024	1.056	.386

The results show that **Income** as one of the variables of demographics factor **has a significant impact on impact Frequency of online purchase** of consumers. Research studies by **Sultan and Henrichs** in 2000 concluded that the consumer's motivation to and inclination for using the Internet as his or her shopping intermediate was positively allied to income the identical is being supported by this study.

#### XI. CONSUMERS' RESPONSE VARIATIONS FOR ONLINE BEHAVIOR \* ACROSS DIFFERENT OCCUPATION GROUPS

(\* Possession of internet, Frequency of online purchase, Motivation drives for online purchase)

Dependent variables	Mean	Std. Deviation	F-Value	Sig.
---------------------	------	----------------	---------	------

Possession of internet	1.11	.320	2.291	.069
Frequency of online purchase	3.41	.577	2.482	.052
Motivation drives for online purchase	2.23	1.024	1.025	.401

The results show that **occupation** as one of the variables of demographics factor does not have any **significant impact on any of the variable taken for study.**

## XII. FINDINGS

The ANOVA results for consumers' response across different demographics factors shows that gender does impact Possession of internet and Frequency of online purchase of consumer's occupation is a demographic variable which does not impact any of the variables under study. The overall results prove that the respondents have perceived online shopping in a positive manner. This clearly justifies the project growth of online shopping. The frequency of online shopping is relatively less in the country. Online shopping organizations can apply the relevant variables and factors, identified from the research, to create their strategies and tactics. The organizations can prioritize the consumer inherent and unequivocal requirements in online shopping environment. The results can also be used by various organizations to identify their target customer segments.

## XIII. CONCLUSION

The results of the study can be utilized by practitioners in relooking or revamping their strategies for online shopping. Online websites should concentrate more to the female segments as results prove that females shop more in online shopping as compared to men. So companies should devise the policies and strategies to magnetize more number of people in this segment in future also. Online retailers should also look into the prospect of call centers which could guarantee that the customer get a chance to officially interact with the other party before the actual purchase. It includes the random sample of individuals from Navi Mumbai. This study finds agreement amongst variety by including people of different age, income, occupation gender segments.

## BIBLIOGRAPHY

- [1] <http://www.livemint.com/Industry/jejcIGVMt1DbB8vKdhvgII/Foreign-retailers-express-concernsover-retail.html> (First Published: Wed, Nov 07 2012.) (07/11/12).
- [2] <http://www.hindu.com/rp/2008/05/02/stories/2008050250130300.htm>.
- [3] [http://articles.timesofindia.indiatimes.com/2012-06-25/news/32408603\\_1\\_retail-sector-retail-industry-global-retail-development-index](http://articles.timesofindia.indiatimes.com/2012-06-25/news/32408603_1_retail-sector-retail-industry-global-retail-development-index) (11/11/2012).
- [4] Fram, E. H., & Grandy, D.B. (1995), "Internet buyers: Will the surfers become buyers?", Direct Marketing, Vol. 57, No. 10, pp. 63-65.

- [5] Fram, E. H., & Grandy, D.B. (1997), "Internet shoppers: Is there a surfer gender gap?", Direct Marketing, Vol. 59, No. 1, pp. 46-50.
- [6] Peterson, R. A., Bal Subramanian, S., & Bronnenberg, B. J. (1997), "Exploring the implications of the Internet for consumer marketing", Journal of Academy of Marketing Science, Vol. 24, No. 4, pp. 329-346.
- [7]. Donthu, N. (1999), 'The internet shopper', Journal of Advertising Research, 39 (3), pp 52-58.
- [8]. Keeney, RL (1999), 'The Value of Internet Commerce to The Customers', Management Science, 45 (4), pp 533-542.
- [9] Bhattacharjee Sarathi Partha, Saha Kumar Anish, Begum Ara Sahin,. 2012 July. The application of E-commerce in Business Application: Their Problems and Prospects. International Journal of Computer Applications (0975 – 8887), Volume 49– No.10. Retrieved from <http://www.euroasiapub.org/IJRESS/Sep2013/4.pdf>
- [10] Zhang Lingying, Tan Wojie, Xu Yingcong, Tan Genlue, Dimensions of Consumers' Perceived Risk and Their Influences on Online Consumers' Purchasing Behavior. CISME Vol. 2, Iss. 7, 2012, PP.8-14, www.jcisme.org ? C 2011-2012, World Academic Publishing.
- [11] Kumari Renu,. (2013) September. Problem and Prospects of E-commerce in Retailing. IJRESS (ISSN 2249-7382), Volume 3, Issue 9. Retrieved from <http://www.euroasiapub.org/IJRESS/Sep2013/4.pdf>
- [12] Patna, H. C. (2013). Is online shopping booming in india? - an empirical study. Retrieved from <http://www.mbaskool.com/business-articles/marketing/7695-is-online-shopping-booming-in-india-an-empirical-study.html>

# ELEMENTAL ANALYSIS OF SOIL SAMPLES USING ENERGY DISPERSIVE X-RAY FLUORESCENCE TECHNIQUE

Vandana Sharma<sup>1</sup>, J.K.Sharma<sup>2</sup>, Shatrugun Olla<sup>3</sup>

<sup>1,2,3</sup> Physics Department Maharishi Markendeshwer University, Mullana, Ambala, Haryana (India)

## ABSTRACT

The work presented in this paper deals with the elemental analysis of soil samples, by using a method based upon Energy Dispersive X-Ray Fluorescence(EDXRF) technique. The soil pollution is a major global problem affecting the physio-chemical properties of the soil. The assessment of pollutant elemental levels and identification of their sources are prerequisite for understanding of their effect on humans, plants and animals. Particularly in the areas near industries there is an urgent need to determine the concentration of heavy metals. The main objective of the present work is to determine the concentration of various elements present in soil in the vicinity and away from industries. In order to check the level of pollutants the determined amounts of elements are compared with their maximum admissible limits. The results of the study show that iron and calcium concentration in soil varies from 8.97 gm/kg - 83.69 gm/kg and from 17.82 gm/kg - 108.0 gm/kg respectively. The present study also reveals that concentration of Fe and Ca is maximum near the industries. However the concentration of these elements decreases as the distance from the industries increases.

**Key words:** EDXRF, Elemental Analysis, Soil pollution.

## I. INTRODUCTION

Soil nutrients are important for plant growth and development. Therefore, the chemical composition of soil, particularly concentration of metal contents is environmentally important. The earth crust contains trace and major elements such as Cs, Cd, Pb, Fe, Mg, Mn, etc. The concentration of all these elements depends on the geology of a local environment as well as on other natural and anthropogenic processes [1]. Small amounts of these elements in soil are necessary for good health, but higher concentrations (more than recommended values) may cause chronic toxicity and disturb the biological balance of organisms in soil and may lead to soil pollution. Soil pollution may be defined as the build-up of persistent toxic compounds, chemicals, salts, radioactive materials, or disease causing agents in soils, which have adverse effects on plant growth and animal health. The most important sources of heavy metals in environment are anthropogenic activities such as mining, steel and iron industry, chemical industry, traffic and domestic activity which may release higher concentrations of these elements relative to normal values and has unfavorable effects on physio-chemical properties of soil [2]. The presence of heavy metals above their recommended values in soil can affect the quality of food, ground water, micro- organism activity and plant growth [3].

The chemicals released by industrial wastes [4], decomposed and partially decomposed chemicals like cadmium, chromium, lead, arsenic, selenium products are likely to be deposited in underground soil and, their accumulation in the soil above their recommendable levels becomes harmful for crops as well as for living

organisms. These can damage the normal activities and ecological balance in the underground soil. Therefore, there is dire need to determine the concentration of heavy elements in the areas nearby the industries when the industrial effluents may affect the concentration of heavy metals in soil.

The aim of the present study is to determine the concentration of heavy metals in soil in the vicinity and within a radius of 1.5Km from the industries using a method (5) based upon EDXRF technique and to compare their concentration with the maximum admissible limits. EDXRF is a rapid, non-destructive multi-elemental analysis technique. The energies of emitted X-rays are the characteristics of elements presents in a sample and thus provide the knowledge about nature of element, hence measurements of energies form the basis of qualitative analysis. Whereas the measurement of the intensities of X-rays provide an information about concentration of elements and thus form the basis of quantitative analysis. Qualitative and quantitative analysis by XRF techniques are performed without digestion and a great number of elements can be determined simultaneously in short time [5-10].

Yamuna Nagar District in Haryana is an industrial town having nearly 2500 small scale and large scale industries. It is second biggest industrial town in Haryana (India) after Faridabad. The main industries in Yamuna Nagar are metal, sugar and paper etc. The industrial waste from these industries may go into the soil and hence can affect the concentration of heavy elements in soil in regions around these industries. Therefore to check the pollutant levels, elemental analysis of soil samples collected from the places close to industries and at a distance 1 to 1.5 Km away from industries had been carried out using a method based upon EDXRF technique[5]. The details of the collection of soil samples, experimental set up, procedure used for elemental analysis, determined amounts and inferences drawn from the study are discussed in this paper.

## **II. SAMPLE COLLECTIONS**

### **2.1 Sampling**

The soil samples from ten different locations of industrial regions of Yumnanagar were collected. Out of ten samples, five samples were taken from close to industries and remaining five samples were taken at distance 1 to 1.5Km away from the industries. The various steps involved in collection of samples and target preparation are discussed below:

Ten soil samples:

- Five from the vicinity of Metal, Wine, Paper, Sugar and ISGEC spare part manufacturing industries,
- And five samples within a radius of 1 to 1.5 km from these industries have been collected.

The metal industry was referred to as sampling site\_1 and the soil samples collected from vicinity and at distance 1 to 1.5 Km from metal industry were referred as Sample\_1(sample close to industry) & Sample\_1D (sample at a distance of 1 - 1.5Km) for the sampling site\_1 in the subsequent discussion. Similarly Wine, Paper, Sugar and ISGEC spare part manufacturing industries, were referred as sampling site\_2, sampling site\_3, sampling site\_4 and sampling site\_5 respectively. The soil samples taken from the vicinity of these sites were labeled as Sample\_2, Sample\_3, Sample\_4, Sample\_5 and samples collected at distance about one Kilometer from the industries were labeled as, Sample\_2D, Sample\_3D, Sample\_4D, Sample\_5D, respectively. Each sample was taken at a depth of 0-10 cm. The samples were collected with a hand auger (a plastic screw) and hand spade and were placed in a clean polyethylene bags to avoid contaminations. The locations of sampling sites are shown in the map of Yamunanagar in Fig. 1.



Fig 1: Map of Yamunanagar Showing The Different Sampling Locations

## 2.2 Traget Preparation

The soil samples were air dried for about a week. All the samples were placed in an oven at 80°C for 3-4 hours. Then samples were grinded in china dish. Each sample was pulverized and homogenized into a very fine powder by passing through 85 micron mesh sieve. Accurately weighed amount of all the soil samples, in the range 3 gm to 5 gm with electronic weighing machine (CPA-225D Sartorius) was pressed in the die of diameter 2.7 cm. The infinitely thick pellets of soil samples were prepared by pressing the die with a hydraulic press (Ankit Hydraulics) and applying pressure in the range 100 to 300 KN. The thicknesses of these pellets were in the range 0.613 to 1.04 in gm/cm<sup>2</sup>.

## III. EXPERIMENTAL SETUP FOR ELEMENTAL ANALYSIS

The elemental analyses of samples were performed using a method based upon EDXRF technique [5]. The experimental setup used for elemental analysis of samples is shown in Fig. 2(a).

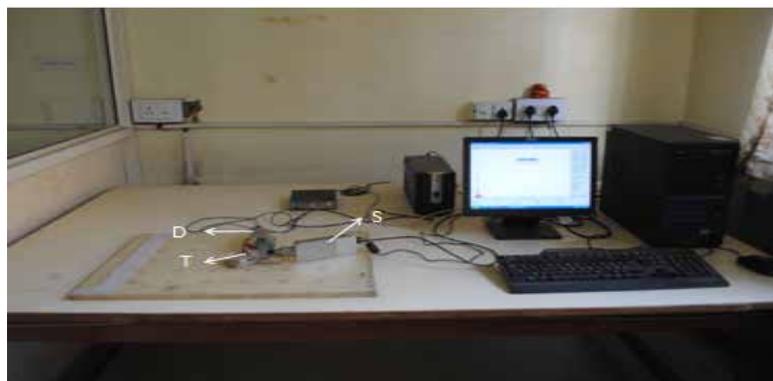


Fig. 2(A): Single Reflection Experimental Arrangement D- Si PIN Detector, S-Source Of Excitation Miniature X-Ray Tube (With Silver (Ag) Transmission Target And A Beryllium Window) And T- Targets Under Investigation.

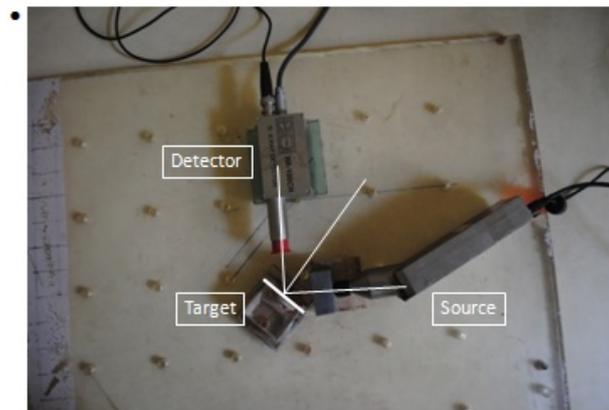
The EDXRF spectrometer consists of

- Self-contained miniature X-ray tube with silver (Ag) transmission target and a beryllium window, a portable controller incorporating the power supply and control electronics. The controller generates all the voltage needed to operate the X-ray tube and provides both voltage (kV) and current (mA) display and control. (purchased from M/s. TENKINS-ELECSYS, LLC, San Diego, USA, AMPTEK MODEL MINI-X-AG)

- X-ray detector Model XR-100 CR(purchased from M/s. TENKINS-ELECSYS, LLC, San Diego, USA) The X-ray detector is ( Model XR-100 CR) high performance thermoelectrically cooled Si-PIN photodiode, with a preamplifier.
- The detector is powered by the PX2 CR power supply. The detector system has a resolution of 150 eV FWHM, for the 5.9 keV peak of <sup>55</sup>Fe. Detector is coupled to MCA8000 A multichannel analyzer for signal processing and data acquisition.

### 3.1 Elemental Analysis of Samples

The elemental analysis of the soil samples were carried out by irradiating sample pellets with 22.16 keV Ag K x-ray in a single reflection geometrical set up (Fig 2(b)).



**Figure 2(b): Single reflection Experimental arrangement of Detector, Source and Targets under investigation.**

The method used in present measurements requires the selective excitation of KX-rays of the element under investigation in

- Sample target under investigation(i.e in soil samples)
- Some known compound of element under investigation referred as first standard
- Sample prepared by mixing known amounts of samples under investigation and first standard referred as second standard

Thick targets of soil samples, first standard and second standard were irradiated with tube photons in single reflection geometrical set up. The intensity of the characteristic X-rays emitted by the elements in soil samples, first standards and second standards were noted from their spectra. These measured values of intensities are used in following relation to calculate the amount of element present in the sample.

$$a = a' \cdot \frac{[N_A(a) / N_{Sa}(a) - 1]}{[N_A(a) / N_{Sa}(a) - N_A(a') / N_{Sa'}(a)]} \dots \dots \dots (1)$$

Where  $a$ : is the amount of element present in sample (to be determined)

$a'$ : is the amount of first standard added to sample to obtain second standard

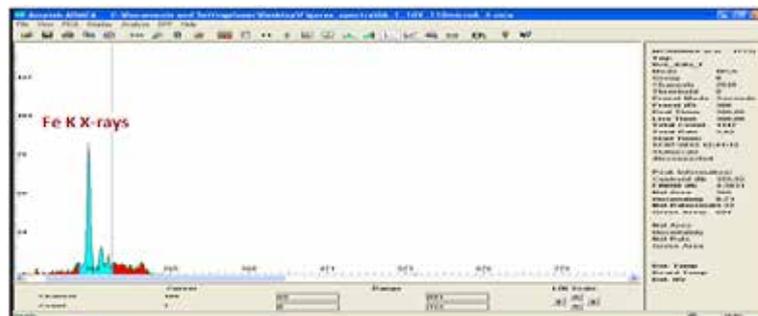
$N_{Sa}(a)$ ,  $N_A(a)$ , &  $N_{Sa'}(a)$ : are the intensities of x rays emitted from sample, first standard and second standard respectively

Each sample pellet was irradiated for 300 seconds at 10 keV and 110 mA. The characteristic x-rays emitted from the targets were recorded with Si PIN detector. Si PIN spectrometer being used in the present

measurements was calibrated using the sources of  $^{55}\text{Fe}$  and  $^{241}\text{Am}$ . The back ground spectra was also recorded by replacing the sample targets with borax target.

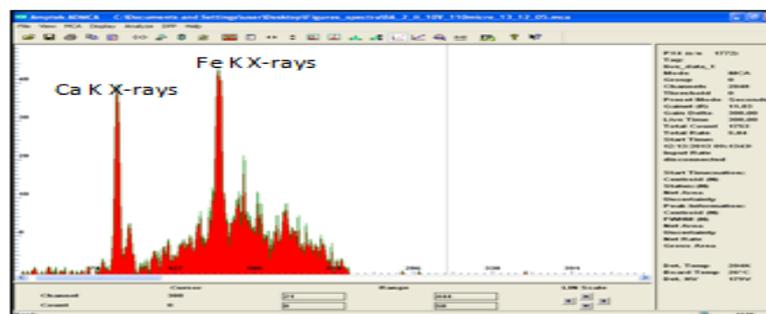
### 3.2 Qualitative Analysis

The qualitative analysis of samples requires the identification of elements present in the sample. It can be carried out quite conveniently by determining the energies of the fluorescent x-rays emitted by elements present in samples. Typical spectra of the radiation emitted from soil samples Sample\_1 and Sample\_2, as recorded by Si PIN Detector were shown in Fig 3(a) and 3(b) respectively. In these spectra, peaks were found at those channels which correspond to the K X-ray energy of calcium and iron. The energies of the peaks appearing in these spectra also correspond to the L X-ray energies of the elements in the  $63 \leq Z \leq 70$ . To check whether the peaks observed in spectra correspond to K X-ray energies of iron/calcium or L X-rays energies of the elements in the range  $63 \leq Z \leq 70$ , the sample pellets were, irradiated with 59.57 keV gamma rays from  $^{241}\text{Am}$ . The resulting spectra were recorded with Si PIN detector and search for K X-ray energies of the elements in the range  $63 \leq Z \leq 70$  was made.



**Fig 3(A): Spectrum of The Radiation Emitted When Soil Sample\_1 Is Irradiated By Ag K X-Rays Emitted By MINI-X-Ray Tube As Recorded By Si PIN Detector In Single Reflection Geometrical Setup**

The net spectra for different samples showed no peaks in the K X-rays energy region of elements  $63 \leq Z \leq 70$ . This rules out the possibility of the presence of L X-rays of the elements with  $63 \leq Z \leq 70$  in the energy range 1—7 keV.



**Fig 3(b): Spectrum of the radiation emitted when soil sample\_2 is irradiated by Ag K X-rays emitted by MINI-X-ray tube as recorded by Si PIN detector in single reflection geometrical set up**

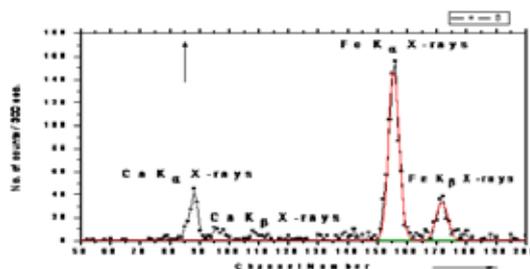
Thus the peaks observed in spectrum correspond to K X-rays energies of calcium and Iron. So, all these observation confirm the presence of Calcium in samples S-2 and S-2D, whereas, Iron in all the samples.

### 3.3 Quantitative Analysis

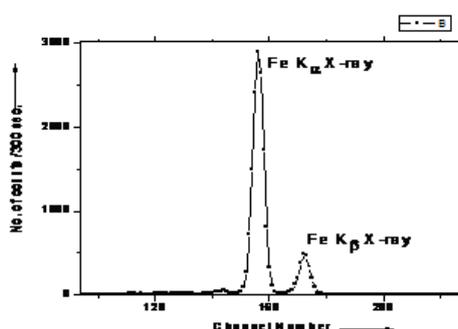
To determine the amounts of iron/calcium in different samples, selective excitation of K X-rays of iron/calcium in samples and its two standards was carried out in single reflection geometrical set up.

#### 3.3.1 Estimation of Iron

For iron determination the pellets of  $\text{Fe}_2\text{O}_3$  were chosen as first standards and known mixtures of sample and  $\text{Fe}_2\text{O}_3$  were used as second standards. The weights of the 1<sup>st</sup> standard materials to be mixed with the sample was estimated from comparison of intensities of characteristics X-rays in the spectra of sample and 1<sup>st</sup> standard, so that the intensity of the emitted X-rays from 2<sup>nd</sup> standard pellet increases considerably. The sample pellets and their two standards were irradiated with tube photons. As the K edge of iron is 7.1 KeV, hence for selective excitation of Fe K X-rays in samples, first standard and second standard the voltage and current of MINI X-tube was kept 10Kv and 110 mA. The resulting spectra of K X-rays were recorded. The recorded spectra also show background. The background spectra with borax were also recorded by irradiating the targets of borax with Ag K X-rays from X-ray tube. The background spectra of borax were subtracted from the spectra of soil samples, first standards and second standards. Typical net spectra of K-X-rays after subtracting the background from soil sample\_2, its first standard and second standards, sample\_2d Sample\_3, are shown in Fig. 4(a-e). The counts under Fe  $K_\alpha$ ,  $K_\beta$  peaks in samples under investigation, their first standards second standards were calculated and used in the relation (1) to determine the concentration of Fe in soil samples. The determined amounts of Fe in soil samples taken from the vicinity and at a distance 1 to 1.5 km were tabulated in Tables 1 & 2 respectively.



**Fig 4(A): Net Spectrum of The Fluorescent KX-Rays From Sample\_2 After Subtracting The Background Spectra of Borax.**



**Fig 4(B): Net Spectrum of The Fluorescent K X-Rays Emitted, When Sample of  $\text{Fe}_2\text{O}_3$ (First Standard For Iron Determination) Is Irradiated With X-Ray Tube (10 Kv & 10microampere) Recorded With Si PIN Detector After Subtracting The Background Spectra Of Borax.**

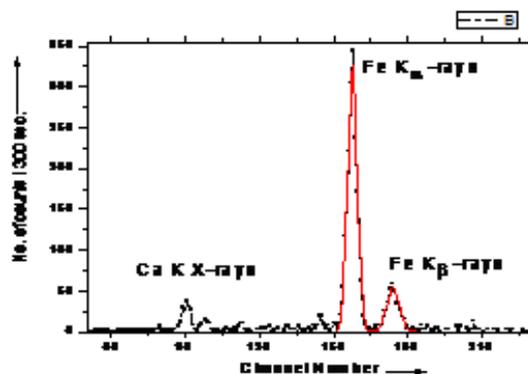


Fig 4(C): Net Spectrum of The Fluorescent K X-Rays From Sample\_2+Fe<sub>2</sub>O<sub>3</sub> After Subtracting The Background Spectra of Borax.

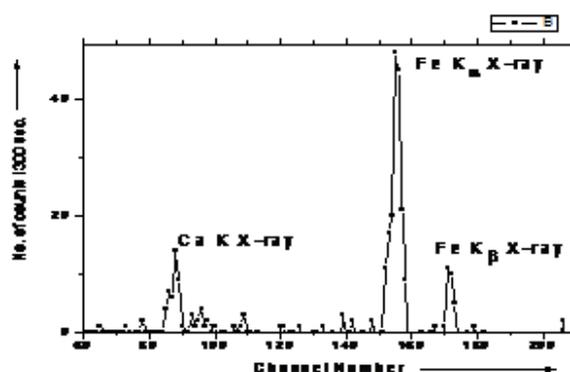


Fig 4(D): Net Spectrum of The Fluorescent K X-Rays From Sample\_2D After Subtracting The Background Spectra of Borax.

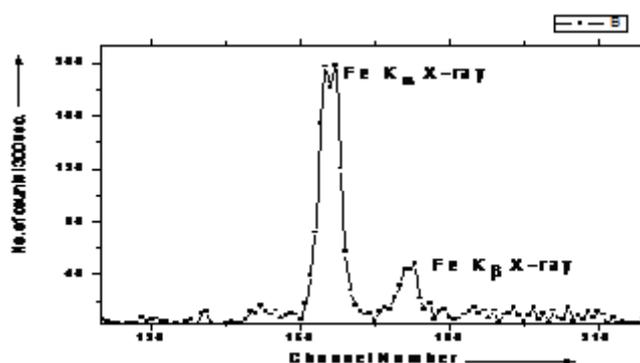
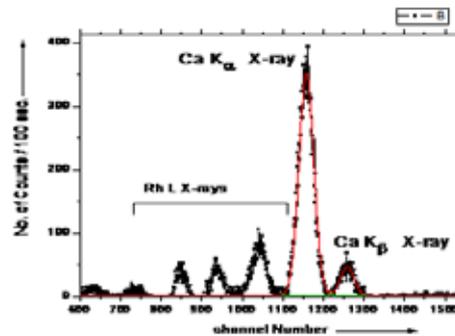


Fig. 4(E): Net Spectrum of The Fluorescent K X-Rays From Sample\_3 After Subtracting The Background Spectra of Borax.

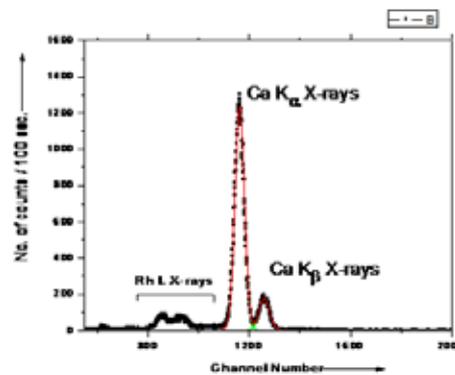
### 3.3.2 Estimation of Calcium

The pellets of CaCO<sub>3</sub> were chosen as first standard for calcium determination and known mixtures of sample & CaCO<sub>3</sub> were used as second standards. In order to determine the amount of Calcium in the presence of Iron, the present method requires selective excitation of Ca K X-rays. The selective excitation of Ca K X-rays can be carried out by keeping the voltage of Mini X-ray tube at about 5V (close to K-edge of Calcium). But the threshold operating voltage of Mini X-ray tube is 10 kV, below 10 kV it does not function. Therefore the determination of Calcium was carried out using Rh K X-rays by keeping the voltage and current at 5kV and 0.2mA. at Nuclear Physics laboratory of Punjabi University Patiala in single reflection geometrical set up. To

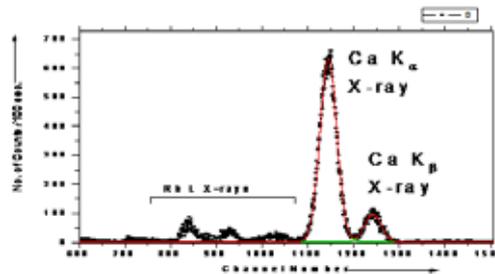
determine the amount of Calcium, the sample pellets and their two standards were irradiated with Rh K X-rays from low power Neptune X-ray tube of Oxford instruments, USA with cooling based Rh target as anode and 2mA / 50 kV. The resulting spectra of K X-rays were recorded with solid state PIN detector (APTEK, USA). The K X-rays spectra for the first and second standard were also taken in the same way by irradiating the targets of first standards and second standards to Rh K X-rays from X-ray tube. The recorded spectra also show background. The background was also recorded with borax. The net spectra of the emitted X-rays from the samples, sample\_2, first standard and second standards after subtracting background spectrum recorded with borax are shown in Fig. 5(a-c). The counts under Ca  $K_{\alpha}$  and  $K_{\beta}$  peaks in samples under investigation, their first standards and second standards were calculated. These counts are used in relation (1) to find the amount of Ca present in samples.



**Fig 5(A): Net Spectrum of The Fluorescent KX-Rays From Sample\_2 After Subtracting The Background Spectra of Borax.**



**Fig 5(B): Netspectrum of The Fluorescent K X-Rays Emitted, When Sample of  $\text{CaCO}_3$  Is Irradiated With X-Ray Tube (5V & 0.2ma) Recorded With Solid State PIN Detector.**



**Fig 5(C): Net Spectrum of The Fluorescent K X-Rays From Sample--2+ $\text{CaCO}_3$ after Subtracting The Background Spectra of Borax.**

**IV. RESULTS AND DISCUSSION**

**4.1 Iron (Fe)**

The determined values of the Fe concentrations in samples taken from the vicinity of industries and at a distance of about 1-1.5 Km away from industries are given in Tables (1) and (2). Present study indicates that Fe is one of the important constituent of soil in Yamuna Nagar region. It is clear from the Tables (1) and (2), that the concentration of Fe for all samples, was found to vary in the range of 8.96 mg/g to 83.69 mg/g. The variation of Fe concentration with the different sampling sites is also shown in Figure(6).

**Table (1):Mass Concentration (In Mg /G) Of Iron (Fe) In Soil Samples Taken From Close To The Industries.**

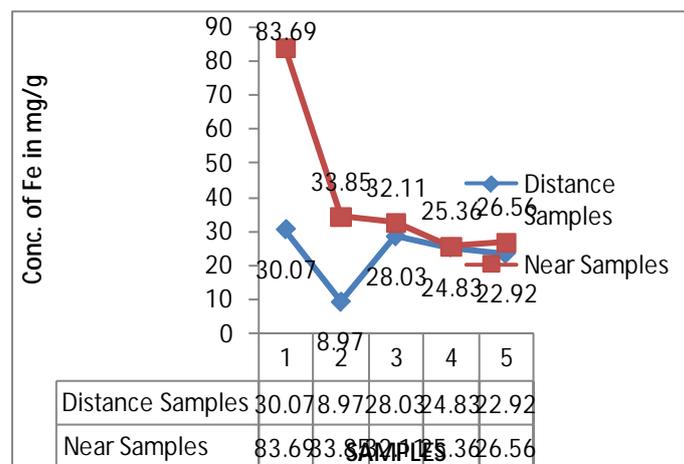
Sr. No.	Samples	Iron (Fe) conc. in mg /g
1.	S 1	83.69
2.	S 2	33.85
3.	S 3	32.11
4.	S 4	25.36
5.	S 5	26.56

Mean value of Fe concentration= 40.31 mg /g

**Table (2) :Mass concentration (in mg /g) of iron (Fe) in soil samples taken from distance (1 to 1.5 km radius) from the main industries.**

Sr. No.	Samples	Iron (Fe) conc. in mg /g
1.	S 1D	30.07
2.	S 2D	8.97
3.	S 3D	28.03
4.	S 4D	24.83
5.	S 5D	22.92

Mean value of Fe conc. =18.15. mg /g



**Fig 6: Comparison of Fe Concentration For Samples Taken From Vicinity & Away From Industries**

Max. Permissible conc. of iron (Fe) according to CCME 1999 =38.00 mg/g [11]

It is observed from Tables (1) and (2) and Fig. 6:

1. The soil samples, sample\_1 and sample\_1D show the highest value of iron content. The concentration of Iron decreases from sample\_1, to sample\_5. The same trend is also observed for the samples collected from the sites which are at distance of 1 to 1.5 Km away from industries.
  2. The measured values of Iron concentrations in different samples were found to be well within the range of recommended value which is 38 mg /g (as recommended by CCME 1999) [11] except for sample\_1. The concentration of Iron found in sample\_1 is 83.69 mg/g which is higher than the recommended value 38 mg/g. The concentration of iron in soil sample collected from a distance of 1km i.e. sample\_1D, is 30.07 mg/g which is below the threshold value.
  3. The concentration of Iron decreases as the distance from the industries increases. The average concentration of iron for samples taken at a distance from the industries is 18.15 mg/g and comparatively lower as compared to the average concentration of iron 40.31mg/gm for samples close to industries.
  4. Also it has been found from figure(6) that the values of Iron concentration in samples (Sample\_4, Sample\_5 and Sample\_4D, Sample\_5D) close to each other.
- Observation (1) i.e Sample\_1 has the maximum values of iron content, can be attributed to the fact that sampling site\_1 is a large scale metal industry and it is near Rampura industrial area. It manufactures heavy spare parts, pipes and makes use of iron as major raw material. Thus, the industrial activity has effect on the concentration of iron on soil in the vicinity of industry.
  - Observation (2) i.e. highest concentration of iron in sample\_1 more than the recommended value indicates the soil in the vicinity of metal industry is contaminated with iron. The determined amounts of Fe in the soil samples taken from sampling site\_2, \_3, \_4, \_5 are well in the range of recommended value, thereby indicating that soil near these industries is not so polluted.
  - Observation (3) i.e. smaller values of Iron content away from the industries indicate that effect of industrial activity on the concentration of element present in soil decreases as the distance from the industries increases. The studies carried by Antoaneta et al [10] also indicate the similar results that heavy metals are present in greater concentration in the vicinity of industrial region as compared to their concentration in soil at a distance away from the industries. The larger values of Iron content close to industries indicate that industrial activity certainly has some effect on the concentration of element present in soil in the vicinity of industries.
  - The sampling site\_4 and sampling site\_5 are close to each other Figure(1) that is why measured iron contents in Sample-4, Sample-5 and Sample-4D, Sample-5D do not show much variation Fig. 6.

## 4.2 Calcium (Ca)

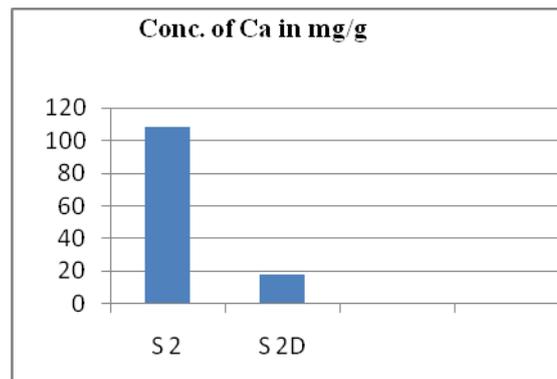
The present study shows that Calcium (Ca) (in measurable concentrations) was present in two soil samples with Sample-2 and Sample-2D. The values of Calcium are found to be 108.500 mg/g and 17.820 mg/g respectively, given in table (3) and Fig. 7. The sampling site-2 from where these soil samples are collected was a small scale wine industry. The raw material used in this industry was fruit e. g. grapes.

The greater amounts of Calcium in sample-2 indicate industrial activity has also some influence on the soil in this case. But smaller concentration of Calcium in Sample- 2D indicates that the effect of industrial activity

decreases with distance. The results indicate that concentration of element decreases with increase in distance from the industry (10).

**Table (3): Mass Concentration (In Mg /G) of (Ca) In Soil Samples**

S.No.	Samples	Calcium (Ca) conc. in mg /g
1.	Sample_2	108.50
2.	Sample_2D	17.82



**Figure (7): Comparison of Ca Concentrations for Samples Taken From Vicinity & Away From Industries**

## V. CONCLUSIONS

Iron and Calcium are most essential minerals that are required for human and plants life for their growth. Small amounts of these elements are necessary for good health, but higher concentrations (more than recommended values) may cause chronic toxicity and disturb the biological balance of organisms in soil. The determined amounts of Iron in the soil samples in the vicinity of industries are greater than those for soil samples taken at a distance away from industries. The determined amounts of Iron in the soil samples in the industrial complexes of Yamuna Nagar are well within the range of recommended values except for one site (Sampling location for sample-1). The determined amounts of Calcium in the soil sample collected from sampling site -2 in the vicinity of industry are greater than those for soil sample taken at a distance away from industries.

Thus conclusions drawn from the present study are:

- The determined amounts of Iron in the soil samples in the vicinity of industries are greater than those for soil samples taken at a distance away from industries.
- The determined amounts of Iron in the soil samples taken at a distance away from industries of Yamuna Nagar are well within the range of recommended values except for one site (Sampling location for sample-1).
- The determined amounts of Calcium in the soil sample collected from sampling site -2 in the vicinity of industry are greater than those for soil sample taken at a distance away from industries.
- The greater values of Fe/Ca contents in the soil samples in the vicinity of industries implies that industrial activities in Industrial region of Yamuna Nagar, do have certain effect on the soil nearby region. However, this effect decreases with increasing distance from the industries.
- Thus the present investigations indicate that soil in the vicinity of metal industry (sampling site-1) is polluted with iron and also soil in the vicinity of wine industry (sampling site-2) shows more value of calcium .However the soil from the regions at a distance 1-1.5 Km. away from the industries is not contaminated.

## VI. ACKNOWLEDGEMENTS

The authors are grateful to Dr. Raj Mittal Singla , Professor Punjabi University Patiala for allowing to use the experimental facilities at Nuclear Science Laboratory, Department of Physics, Punjabi University Patiala.

## REFERENCES

- [1] Mihaela Ulmanu, Ildiko Anger, Eugenia Gament, M. Mihalache, Georgiana Plopeanu, L. Ilie, Rapid determination of some heavy metals in soil using an X-ray fluorescence portable instrument, Research Journal of Agricultural Science, 43 (3) ( 2011)235.
- [2] Tripathi, D. R. Misra, A study of physico-chemical properties and heavy metals in contaminated soils of municipal waste dumpsites at Allahabad, India, International journal of environmental sciences 2, (2012).
- [3] Koranteng Addo , Owusu Ansah, L.K. Boamponsem, J. K Bentumand J. K S. Arthur, Levels of zinc, copper, iron and manganese in soils of abandoned mine pits around the Tarkwa gold mining area of Ghana, Advances in Applied Science Research, 2 (1), (2011): 280-288.
- [4] A. Ene, A. Bosneaga, L. Georgescu, Determination of heavy metals in soils using XRF technique Rom. Journ. Phys. 55 (2010) 815-820.
- [5] Vandana and Raj Mittal, Effect of calcium and potassium fertilizers on calcium and potassium contents of rice saplings, Ecol. Env. & Cons. 5(4) (1999) 373-380.
- [6] Kanan Deep, Preeti Rao, Himani Bansal, Raj Mittal, Matrix effects for calcium and potassium K-X-rays, in fenugreek plants grown in iron rich soils, Appl. Radiation& Isotopes, 90C (2014) 109-116.B.
- [7] Kanan Deep, Raj Mittal, Macronutrient k variation in mung bean sprouts with lunar phases, European Scientific Journal 10(9) (2014) 295-306.
- [8] Michael Mantier, Manfred Schreine, X-ray fluorescence spectrometry in art and archaeology, X-Ray Specrom. 29 (2000) 3-17.
- [9] Antoaneta Ene, Alina Bosneaga, L. Georgescu, Determination of heavy metals in soils using XRF technique, Rom. Journ. Phys, 55 (2010) 815-820.
- [10] Antoaneta Ene, Ana Pantelica, Carmofreitas, Alina Bosneaga, EDXRF and INAA Analysis of soils in the vicinity of a metallurgical plant, Rom. Journ. Phys, 56 (2011) 993-1000.
- [11] Vandana Partha, N.N. Murthy and Parveen Raj Sexena, Assessment of heavy metal contamination in soil around hazardous waste disposal sites in Hyderabad city (India): natural and anthropogenic implications, E3 Journal of Environment and Management 2(2) (2011) 027-034.

# ASSESSMENT OF USE OF RICE HUSK ASH IN SELF CONSOLIDATING CONCRETE

**Mr. Sumit Ahlawat<sup>1</sup>, Mr. Utsav Jain<sup>2</sup>, Mr. Shwetang Kundu<sup>3</sup>**

*<sup>1,2,3</sup> UG Student, Department of Civil Engineering, ITM University Gurgaon, (India)*

## ABSTRACT

*The main objective of this study is to investigate the utilization potential of Rice Husk Ash (RHA) in self-consolidating concrete (SCC). The effect of Ordinary Portland Cement (OPC) partial substitution with RHA on the mechanical properties such as workability, compressive strength, and flexural strength of the SCC were investigated. RHA which is an Agricultural Waste and is used because of its super pozzolanic properties is used for production of SCC. The tests were conducted to find the optimal mix design for best compressive strength. Further investigations were carried out to determine the flexural properties of SCC to obtain the moduli of rupture and tensile strength. The values obtained from the investigations are 7.56 MPa, 9.82 MPa and 6.57 MPa for tensile strength for 5 %, 10 % and 15 % respectively. The tests are conducted to determine the strength characteristics of the SCC, when cement partially substituted in different proportions of RHA. The preliminary results of research based on partial replacement of cement with RHA in SCC and its effectiveness in analysed and computed.*

**Keywords: Ordinary Portland Cement (OPC), Selfconsolidating Concrete (SCC), Rice husk ash (RHA)**

## I. INTRODUCTION

India being an Agriculture based economy, produces a considerable amount of by-products such as rice husk. In this world fighting with the enemies like global warming and shortage of natural construction materials, we are entitled to find alternatives which minimise the environment degradation as well as provide comparable output to the traditional materials. We performed experimentations to determine and evaluate the effectiveness of using RHA in production of (SCC). As RHA is produced in huge quantities at power plants across the country, it is readily available and lowers down the pressure on cement industry because it can be used as a partial replacement to cement. Moreover it reduces the greenhouse gases produced during the production of Cement and as well reduces the need to dump the Rice husk Ash generated.

Self-consolidating concrete (SCC) is concrete that is characterised by low yield stress, high deformity and moderate viscosity that is just sufficient to ensure the suspension of the solid particles during transportation, placement and until the concrete hardens. Because of its flow ability it shows remarkable results in heavily reinforced structural members.

RHA is obtained on burning an Agro-waste, Rice Husk under controlled temperature condition with temperature below 800 °C, the process reduces the Rice Husk into ash which is nearly 25% by weight. It is a super pozzolanic material because of the presence of silica (85% to 90%) and alumina (5%). pozzolanic material can be defined as siliceous or siliceous & alumina substance, having very less or absolutely nocementious property

but possessing cementite properties, when present in fine particle form in presence of moisture react with calcium hydroxide at room temperature.

We have focussed our experimentation on mechanical properties like workability, compressive and flexural strengths of SCC to determine viability of RHA substitution.

## II. MATERIAL

### 2.1 Rice Husk Ash (RHA)

Rice Husk which is burnt for 36hrs (approx) under uncontrolled burning in the an open area. The temperature being in the range of 450- 550 degree Celsius. The ash so generated was sieved by a sieve of size 75 $\mu$ m (BS standard sieve) and the RHA hence collected was of grey colour. Batching for the same is done by volume method.

### 2.2 Cement

Ordinary Portland cement (OPC) grade 43 is being used. (accordance IS 8112:2013)

### 2.3 Aggregates

The aggregate used is such that it passes through 16mm sieve and is retained on 10mm sieve. The Percentage water absorbed when dipped in water for 24 hours was 0.75 % by weight which is in compliance with IS 2386 – Part 3: 1963. Proper Inspection was carried to ensure absence of deleterious materials in aggregate used.

### 2.4 Water

The water used was tap water and was free from any kind of contaminants such as oils, acids, salts, sugars or organic materials. The water had a pH of 8.

### 2.5 Super Plasticizer

The super plasticizer used is SS-PLAST-PC-100. SS-PLAST -PC-100 is new generation super plasticizer for long retention. It is formulated with synthetic polymers with High Molecular weight poly carboxylate and organic polymers. SS-PLAST-PC-100 conforms to IS: 9103-1999, IS: 2645-1975 and ASTM C-494 type F & G

**Table 1: Technical Properties of Super Plasticizer**

Colour	Light Yellowish
Form	Viscous Liquid
Specific Gravity/ Density	1.10±0.02
Chloride	Nil
pH	6 to 7
Air Entertainment	Max. 1.5% of control as per IS 9103: 1999
Water Reduction	More than 35%

**Table 2: Chemical and Physical Properties of RHA (Wt %)**

Chemical properties of RHA									
SiO <sub>2</sub>	Al <sub>2</sub> O <sub>3</sub>	Fe <sub>2</sub> O <sub>3</sub>	CaO	MgO	SO <sub>3</sub>	Na <sub>2</sub> O <sub>3</sub>	K <sub>2</sub> O	LOI	Reference
92.5a	1.2	2.1	0.9	0.4	0.1	2.0	0	0.9	
88.5b	1.1	2	0.9	0.4	0.1	2.6	0	0.9	[1]
87.32	0.22	0.28	0.48	0.28	-	1.02	3.14	2.1	[2]
87.6	0.68	0.93	1.3	0.35	-	0.12	2.37	-	[3]
93.2	0.4	0.1	1.1	0.1	0.9	0.1	1.3	3.7	[4]
80	3.93	0.41	3.82	0.25	0.78	0.67	1.45	8.65	[5]
87.2	0.15	0.16	0.55	0.35	0.32	1.12	3.6	6.55	[6]

**Chemical Properties of OPC**

20.25	5.04	3.16	63.61	4.56	-	0.08	0.51	3.12	[2]
21.89	5.3	3.34	53.27	6.45	3.67	0.18	0.98	3.21	[3]
20.9	4.8	3.4	65.4	1.3	2.7	0.2	0.4	0.9	[4]

**Physical properties of RHA**

Material specific gravity(gm/cm <sup>3</sup> )	Mean particle size (Um)	Blaine's Fineness	References
2.3a, 2.27b	-	4750(cm <sup>2</sup> /gm), 5750(cm <sup>2</sup> /gm)	[1]
2.06	2.5	36.47(m <sup>2</sup> /kg)	[2]
2.23	10.00	11200(cm <sup>2</sup> /gm)	[4]
2.11	-	-	[5]
2.06	8	28800(m <sup>2</sup> /kg) nitrogen absorption	[6]
3.11	-	3250(cm <sup>2</sup> /gm)	[1]
3.14	14.6	-	[7]
3.14	15	3600(cm <sup>2</sup> /gm)	[8]
3.1	22.5	326(m <sup>2</sup> /kg)	[2]

a,b – RHA obtained from electric power station and rice mill respectively

**III. METHODOLOGY**

SCC is a concrete consisting of cement, aggregate, water and chemical admixtures. It is essential that the material remains in suspension and uniformly distributed throughout the process of handling, placing and until setting.

The following tests were carried to determine the workability:

**3.1 Slump Flow and T500 Test**

Slump flow test and T500 time as the test performed to determine the flow ability and the flow rate of SCC, when it is allowed to flow freely without any obstructions. The tests performed were in compliance with the procedure described in EN 12350-2[9]. The slump flow results depict the filling ability of SCC, T500 is the speed of flow and hence help to determine the viscosity of SCC.

**3.2 V- Funnel Test**

The V- Funnel test is used to determine the filling ability and viscosity of SCC. The test is performed in accordance to EN 12350-1[9].

**3.3 L-Box Test**

The L-Box test is being used to determine the passing ability of SCC, the passing ability plays a major role in movement of concrete through narrow opening including the space between reinforcement bars, without blocking and requiring vibrations. The test was performed on three bar apparatus. The three bar apparatus is used when the structure is heavily reinforced. The test was performed in accordance with EN 12350-1[9].

## IV. RESULTS AND DISCUSSION

The experimental results for RHA, cement, and cement SCC is determined and listed below:

### 4.1 Rice Husk Ash

The following values were obtained after performing experiments as per Indian Standard codes

- Ø Normal Consistency = 19%
- Ø Initial Setting time = 200 minutes
- Ø Final setting time = 290 minutes
- Ø Compressive strength = 11N/mm<sup>2</sup>
- Ø Specific gravity = 2.09

### 4.2 Ordinary Portland Cement

- Ø Normal Consistency = 24 %
- Ø Initial setting time = 30 minutes
- Ø Final setting time = 10 hours
- Ø Specific Gravity = 3.1

### 4.3 Fresh Self-Consolidating Concrete

The following tests were performed to determine the workability of fresh concrete and hence average value are listed below:

Slump Flow test -693mm

V Funnel Test -2.8s

L box Test-0.8

## V. MIX PROPORTIONING

The mix proportioning is done in compliance to EFNARC and Indian standard recommendations in IS 10262-1982. The mean target strength was 31.6 N/mm<sup>2</sup> for the Mix proportion. The total binder material was 316.36 kg/m<sup>3</sup>, fine aggregate was taken to be 903kg/m<sup>3</sup> and coarse aggregate was taken 739 kg/m<sup>3</sup>, the water binder ratio was kept 0.45. The super plasticiser content was kept 1.5% of the total cementing material. The mixing was carried out in a pan mixer for 5 minutes, then the moulds were filled and kept for 24 hours before de-moulding. The moulds were kept in curing tank until the day of testing. The moisture content and water absorption of the aggregate and Rice husk ash were taken into consideration and hence the accordingly. The cubes were kept in curing tank for 7, 28 and 48 days. The cement was replaced in percentages of 0,5,10,15,20,25 and 30 with RHA and 150x150x150mm cubes were used for casting. The compaction was provided in three layers with 25 number of blows water content corrections were applied.

## VI. TEST METHODS

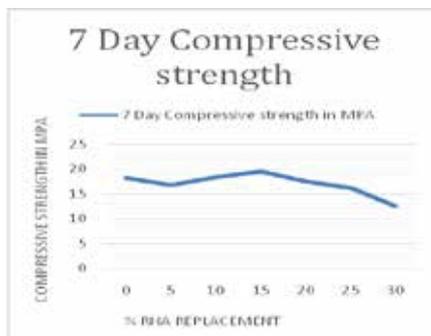
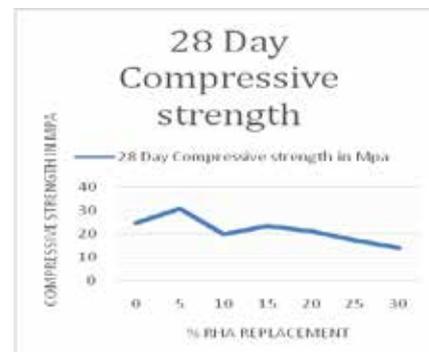
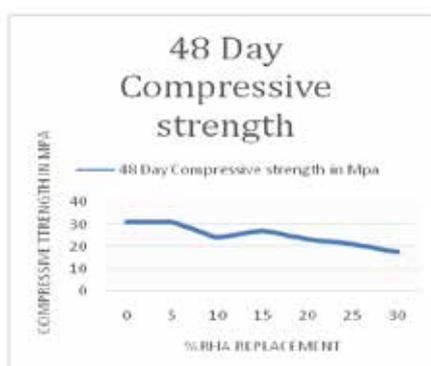
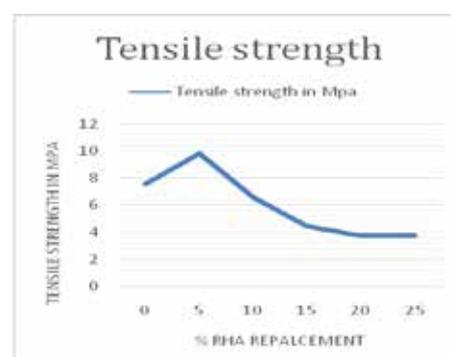
The compressive strength testing of cubes is done according to the IS 516– 1959.

**Table 3: Workability test Values**

S no.	Mix		Workability		
	Cement (%)	Rice Husk Ash (%)	Slump Test	L- Box Test $P_L$	V- Funnel Test
1	100	0	695	0.61	2.2
2	95	5	663	0.66	2.8
3	90	10	694	0.64	2.2
4	85	15	657	0.666	2.8
5	80	20	673	0.67	2.7
6	75	25	701	0.65	2.1
7	70	30	695	0.71	2.8

**Table 4: Compressive strength results**

S no.	Mix		Strength after curing in Days in N/mm <sup>2</sup>			
	Cement (%)	Rice Husk Ash (%)	7 Days	28 Days	48 Days	Tensile Strength
1	100	0	18.11	24.65	30.96	7.56
2	95	5	16.67	30.82	30.821	9.82
3	90	10	18.26	19.78	24.04	6.57
4	85	15	19.47	23.43	27.09	4.43
5	80	20	17.43	21	23.13	3.75
6	75	25	16.13	17.35	21.06	3.75
7	70	30	12.55	14	17.65	2.46

**Figure 1: Compressive Strength 7 Day****Figure 2: Compressive Strength 28 Day****Figure 3: Compressive Strength 48 Day****Figure 4: Tensile Strength with different mix Proportion**

## VII. CONCLUSIONS

Based upon the experimentation results above, the following conclusions are made:-

- 1) Water cement ratio is increased with addition of RHA as RHA is a highly porous material.
- 2) With increase in replacement of RHA, the workability of SCC decreases.
- 3) RHA being rich in silica is preferred over silica fumes to increase the strength.
- 4) RHA particles being very fine, it is harmful for human being, but the cost of RHA is zero. Thus we prefer RHA over silica fumes.
- 5) The compressive strength increases with an increase in percentage of RHA up to 5% of cement in SCC.
- 6) RHA SCC workability can be increased with use of chemical admixtures such as plasticiser, super plasticiser.
- 7) Due to high specific surface area of the RHA, the content of super-plasticiser was increased in Design mix [1].
- 8) RHA addition in the concrete evidently reduces the weight of SCC after 90 Days of moist curing.

## ACKNOWLEDGEMENT

We would like to thank Ms. PrachiSohoni (Assistant Professor, Department of Civil and Environmental Engineering, ITM University, Gurgaon ), for her continuous support and encouragement. We express our sincere gratitude towards Department of Civil and Environmental Engineering, ITM University, Gurgaon (HR) for supporting us and providing all necessary laboratory equipments.

## REFERENCES

- [1] Chatveera, B. and P. Lertwattanaruk, 2009. "Evaluation of Sulfate Resistance of Cement Mortars Containing Black Rice Husk Ash. Journal of Environmental Management, 90(3): 1435-1441.
- [2]Ganesan, K., K. Rajagopal and K. Thangavel, 2008 " Rice Husk Ash Blended Cement: Assessment of Optimal Level of Replacement for Strength and Permeability Properties of Concrete" Construction and Building Materials, 22(8): 1675-1683.
- [3] Givi, N.A., S.A. Rashid, F.N.A. Aziz and S.M.A. Mohd, 2010. "Assessment of the Effects of Rice Husk Ash Particle Size on Strength, Water Permeability and Workability of Binary Blended Concrete" Construction and Building Materials, 24(11): 2145-2150.
- [4] Chindaprasirt, P. and S. Rukzon, 2008. "Strength, Porosity and Corrosion Resistance of Ternary Blend Portland Cement, Rice Husk Ash and Fly Ash Mortar" Construction and Building Materials, 22(8): 1601-1606.
- [5]Ismail, M.S. and A.M. Waliuddin, 1996."Effect of Rice Husk Ash on High Strength Concrete. Construction and Building Materials", 10(7): 521-526.
- [6] De Sensale, G.R., 2006. "Strength development of Concrete with Rice-Husk Ash. Cement and Concrete Composites", 28(2): 158-160

- [7] Tangchirapat, W., C. Jaturapitakkul and P. Chindapasirt, 2009. Use of Palm Oil Fuel Ash as a Supplementary Cementitious Material for Producing High-Strength Concrete. Construction and Building Innovation; Fundamental Research Grant Scheme (FRGS); Materials, 23(7): 2641-2646
- [8] Rukzon, S. and P. Chindapasirt, 2009b. "An Experimental Investigation of the Carbonation of Blended Portland Cement Palm Oil Fuel Ash Mortar in an Indoor Environment". Indoor and Built Environment, 18(4): 313-318.
- [9] The European Project Group (2005) "The European Guidelines for Self Consolidating Concrete". SCC European Project Group.
- [10] Okamura, H.; and Ouchi, M. (2003)" Self-consolidating concrete. Journal of Advanced Concrete Technology", 1(1),Pg 5-15.
- [11] Salas, J. (1988). "Crucial curing of rice husk concrete" Build. Res. Pract., 21(6), 367-376.
- [12] IS: 383-1970. Specifications for coarse and fine aggregates from natural sources for concrete. New Delhi, India: Bureau of Indian Standards.
- [13] James, J., and Rao, M. S. P. (1986). "Reaction products of lime and silica from rice husk ash." Cem. Concr. Res., 16(1), 67-73.
- [14] MuhammadHarunur Rashid, Md. Keramat Ali Molla and Tarif Uddin Ahmed. "Mortar Incorporating Rice Husk Ash: Strength and Porosity", Khulna University of Engineering & Technology, Bangladesh ISSN 1450-216X Vol.40, No.3 (2010), pp.471-477.
- [15] N. R. Gaywala, D. B. Raijiwala, Self-consolidating Concrete: A concrete of next decade, Journal of Engineering Research and Studies, 2(4), 2011, 213-218.
- [16]. A. Juma, E. Rama Sai, et al., A Review on Experimental Behavior of Self Compaction Concrete Incorporated with Rice Husk Ash, International Journal of Science and Advanced Technology, 2(3), 2012, 75-80.
- [17]. Safiuddin, J. S. West, K. A. Soudki, Hardened properties of self-consolidating high performance concrete including rice husk ash, Cement and Concrete Composites, 32(9), 2010, 708-717.
- [18] IS: 8112- 1989. Specifications for 43 grade ordinary Portland cement. New Delhi, India: Bureau of Indian Standards

# A SURVEY ON LOAD BALANCING ALGORITHMS IN CLOUD COMPUTING

Harshada Raut<sup>1</sup>, Kumud Wasnik<sup>2</sup>

<sup>1</sup>M.Tech. Student, Dept. of Computer Science and Tech., UMIT, S.N.D.T. Women's University, (India)

<sup>2</sup>Professor, UMIT, S.N.D.T. Women's University, Mumbai, Maharashtra, (India)

## ABSTRACT

Cloud computing is sharing various computing resources rather than having local servers or personal devices to handle applications over internet. There is lots of data stored on cloud and multiple requests come for resources, due to this increase load on data servers and sometimes unable to provide resources on time. So to manage the load by scheduling task appropriately researchers have introduced some load balancing techniques. Job scheduling can be done by various algorithms like FCFS, Round Robin, and Priority Scheduling. In this paper we are reviewing existing static and dynamic load balancing techniques like Min-Min, Max-Min, Ant colony optimization etc.

**Keywords:** Cloud Computing, Dynamic Load Balancing, Load Balancing, Load Balancing Algorithms, Static Load Balancing, Task Scheduling.

## I INTRODUCTION

Recently cloud computing has become most popular field for academic research area. Cloud computing is a central remote servers where a programs or applications run on server machine. Cloud computing is a system where services like IaaS (Infrastructure is provided as a Service), PaaS (Platform is provided as a Service), SaaS (Software is provided as a Service) are provided online using Internet Connection. These services are provided anywhere, anytime on the basis of "only pay for what you use".



Fig 1: Cloud Computing System

Cloud Computing consists of Virtual machine, Host, Servers. User need not to worry about maintenance of any software, it is the responsibility of cloud service provider to provide the needed resources to user. On the basis of package the user gets the services from service provider, but to use cloud services user needs continuous internet connection. It is the responsibility of cloud Service provider to provide resources to user as per user's demand. Fig.1 shows cloud computing system where cloud services SaaS, PaaS, IaaS are provided to uses.

The cloud appears to be a single point of access for all the computing needs of consumers. It is difficult to manually assign tasks to computing resources in clouds because hundreds of thousands of virtual machines (VMs) are used [1]. Service provider should provide all cloud services to user efficiently, but sometimes due to too many requests from users to utilise resources machine may get overloaded. In results, it is unable to provide resources to users and performance falls down. To avoid this, a good scheduling algorithm is required. Scheduling algorithm helps to provide resources to all users, use all machines and make sure there should not be under loading or overloading machines.

There are 4 Types of Cloud Deployment Models:

Ø *Private Cloud*

This model is build where single organisation wants its own cloud infrastructure for private use. This may be managed by internally or by a third Party.

Ø *Public Cloud*

In this type of model a cloud infrastructure is provided to all users with less access restrictions. Service is easily available and free to all users so it is less secure.

Ø *Community Cloud*

This infrastructure is used by some specific organizations which are in the same group. This infrastructure may be managed by internally or by a third party.

Ø *Hybrid Cloud*

Hybrid is consisting of more than 1 cloud which may be combination of Private, Public or Community Cloud. These clouds would typically be created by the enterprise and management responsibilities would be split between the enterprise and cloud provider [2].

## II LOAD BALANCING IN CLOUD COMPUTING

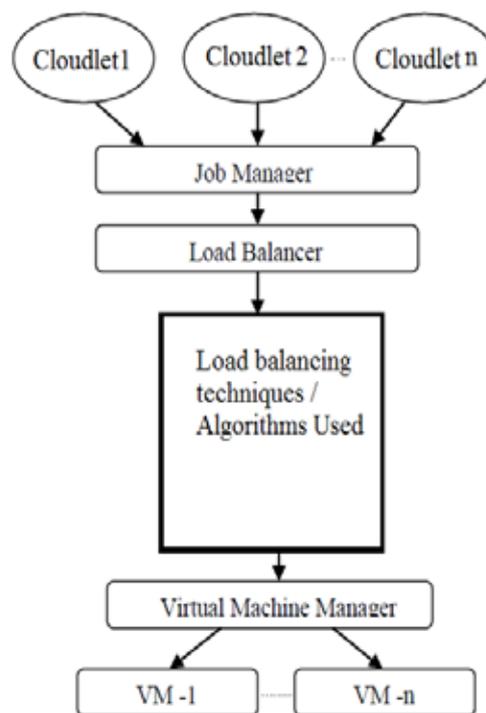
In cloud computing load balancing is a major issue. As the time passes number of users may get increase and this will also increase the load on server and decrease the performance of functions provided by cloud service provider. If there are multiple requests at a same time from different users for resource utilization then it create a long queue on server and increases a response time and to avoid this, divide the task among all available machines and balance a system. If resources are less compared to requests then it may delay the response to clients and server may not be able to fulfil all demands of clients.

Load Balancing is dividing total amount of work in all available machines for execution. To balance a load, each task should be scheduled very carefully so every user may get service in less waiting time. There are various algorithms used for load balancing like Round robin, Ant colony optimization, Particle Swarm Optimization, Max-Min, Min-Min etc.

The aim of load balancing is as follows:

- Quick response time from service provider
- The availability of resources in less waiting time
- No system failure for large task
- Provide the service to every user
- Improve performance
- Accommodate future modification[3]

Fig 2. shows a framework of under which load balancing algorithms work in a cloud computing environment. Cloudlet submits tasks to Job Manager and then Job Manager gives all jobs to Load Balancer. Load Balancer applies Load Balancing Algorithm for Submitted Tasks and schedule all task such that each Virtual Machine will get equal number of task for execution.



**Fig 2: Load Balancing Algorithms Execution [4]**

## **2.1 Load balancing can be done by two types**

### **2.1.1 Static load balancing**

Load is balanced by dividing the tasks before starting of execution in machine. In static load balancing, Load dividing decision does not depend upon the current state of machine.

Round Robin, Opportunistic Load Balancing, Min-Min, and Max-Min are the static load balancing algorithms.

### **2.1.2 Dynamic load balancing**

Load is balanced by considering the current state of the machine and then decision is made. It checks the network and assign task to machine which is underloaded or ideal.

Ant Colony Optimization, Honeybee Foraging, Biased Random Sampling, Active Clustering are the Dynamic load balancing algorithms.

## **II EXISTING LOAD BALANCING ALGORITHMS**

In this section we discuss various load balancing algorithms which are already implemented and are available for load balancing.

Ant colony optimization (ACO) algorithm is proposed by Marco Dorigo in 1992. It is inspired from real ants, when searching for a food ant travels randomly and in return trip they deposited some chemical pheromone. On the quantity of this pheromone other ants uses shortest path on which more pheromone value is deposited. This algorithm helps to find a minimum distance from source to destination. In [5] enhancement to ACO algorithm is proposed. Artificial ants moves forward and backward direction to find the overloaded node and update value in pheromone table. Throughput, Response time, less energy consumption is achieved, but it gives less performance.

Honeybee Foraging Behaviour Algorithm is dynamic approach. [6]Baris Yuce et. al proposed enhanced version of Bees algorithm called ANSSA. It is a decentralized honeybee-based load balancing technique. It is a nature-inspired algorithm for self-organization. In this algorithm sends bees to selected sites and evaluates a value which is best suited for that site. It gives better result for higher dimensions in local server actions. System Performance is improved with increased system diversity but throughput is not increased with an increase in system size. It is suitable for the conditions where the diverse population of service types is required.

Round Robin is a static behaviour algorithm. It uses time quantum and allocate jobs in first come first serve manner, after allocating jobs to all machine sequentially, it return to first machine and allocate next job if that machine is finishes its time quantum, jobs are not allocated priority wise. In paper [7], Nusrat Pasha et.al proposed enhancement to round robin algorithm. In this, process does not wait for total consumption of resources it move to next service. This achieves resource utilization and better response time. Sometimes more load on server if more processes trying to switch in short time.

Min Min[8] is a static load balancing algorithm. This algorithm selects job with minimum execution time and assigned it to machine which can executes that job in minimum time. This algorithm is suited where small task are more compared to large task as it executes smaller task first and gives less priority to large task in execution. More waiting time for large size task.

Max Min is static load balancing algorithm, this algorithm is overcomes the disadvantage of Min-Min algorithm. In this, select jobs with maximum execution time and assigned it to machine with min completion time [9]. In other words, larger jobs run on slower machine. The enhancement[10] to this Max-min Algorithm is instead of selecting maximum execution time task, selects an Average or nearest greater than average task then overall makespan is reduced and also balance load across resources.

**III COMPARISON TABLE OF EXISTING LOAD BALANCING ALGORITHMS**

Algorithms works in Environment	Algorithms	Resource Utilization	Throughput	Response Time	Waiting Time
Dynamic	ACO	Yes	Yes	Yes	No
Dynamic	Honeybee Foraging	No	Yes	No	No
Static	Round Robin	Yes	Yes	Yes	No
Static	Min Min	Yes	Yes	Yes	Yes
Static	Max Min	Yes	Yes	Yes	No

**IV CONCLUSION**

In this paper, we discussed and compared some load balancing algorithms. To balance a load on vast cloud system a proper scheduling of task is required. For satisfactory service a user needs flexibility in service, easy availability, less waiting time to get resource etc. Incoming jobs can be scheduled according to virtual machine's capacity so that it may avoid the unavailability of service if there are multiple requests on single machine and this will get satisfactory service to cloud users. There is a necessity of permanent solution for load balancing in cloud computing. All algorithms discussed here have some advantages and disadvantages.

**REFERENCES**

- [1] Qiyi, H., Tinglei, H., "An Optimistic Job Scheduling Strategy based on QoS for Cloud Computing", IEEE International Conference on Intelligent Computing and Integrated Systems (ICISS), pp.673-675, 2010.
- [2] Rabi Prasad Padhy, Manas Ranjan Patra, Suresh Chandra Satapathy, "Cloud Computing: Security Issues and Research Challenges", *International Journal of Computer Science and Information Technology & Security (IJCSITS)*, Vol. 1, No. 2, December 2011
- [3] Dharmesh Kashyap, Jaydeep Viradiya, "A Survey Of Various Load Balancing Algorithms In Cloud Computing", *International Journal of Scientific and Technology Research*, Volume 3, Issue 11, 2014
- [4] N. S. Raghava, Deepti Singh "Comparative Study on Load Balancing Techniques in Cloud Computing", *Open Journal of Mobile Computing and Cloud Computing*, Volume 1, Number 1, August 2014.
- [5] Shagufta Khan, Nireesh Sharma, "Effective Scheduling Algorithm for Load balancing (SALB) using Ant Colony Optimization in Cloud Computing", *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 4, Issue 2, February 2014
- [6] Baris Yuce, Michael S. Packianather, Ernesto Mastrocinque, Duc Truong Pham and Alfredo Lambiase "Honey Bees Inspired Optimization Method: The Bees Algorithm", *Insects* 1 July 2013; Published: 6 November 2013.

- [7] Pooja Samal, Pranati Mishra, "Analysis of variants in Round Robin Algorithms for load balancing in Cloud Computing", *International Journal of Computer Science and Information Technologies*, Vol. 4 (3) , 2013, 416-419
- [8] T. Kokilavani, Dr. D. I. George Amalarethnam, "Load Balanced Min-Min Algorithm for Static Meta Task Scheduling in Grid computing", *International Journal of Computer Applications Vol-20 No.2*, 2011
- [9] Santhosh, Dr. Manjaiah D.H, "An Improved Task Scheduling Algorithm based on Max-min for Cloud Computing", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol.2, Special Issue 2, May 2014
- [10] Upendra Bhoi, Purvi N. Ramanuj, "Enhanced Max-min Task Scheduling Algorithm in Cloud Computing", *International Journal of Application or Innovation in Engineering & Management (IJAIEM)*, Volume 2, Issue 4, April 2013.

# INTER-CLUSTER COMMUNICATION AND INTRUSION DETECTION FOR EFFICIENT DATA TRANSMISSION IN WIRELESS SENSOR NETWORKS

M.S.Vijayakumar<sup>1</sup>, Dr.S.V.Manisekaran<sup>2</sup>

<sup>1</sup> PG Scholar, <sup>2</sup> Assistant Professor, Department of IT,  
Anna University Regional Centre Coimbatore, Tamilnadu (India).

## ABSTRACT

*In wireless sensor network, during the process of communication the information is lost due to lack of security in the network. Several authentication methods are present to increase the security in the wireless sensor network. But still many Active and passive attacks are present in the network to reduce the privacy of the network. The current state of art in the authentication protocols will take us to the conclusion that the topic is still open. So in the proposal using the clustering, intrusion detection and fusion methods to overcome the active and the passive attacks in the network. PEGASIS Clustering is used to improve the efficiency of energy. The Algorithm is used for intrusion detection is Rational Operating recall Curve Algorithm. It works to reduce the intrusion rate of the system during the process of communication in the network. The fusion method is also introduced to find the precision rate of the network. By using this method the security of the wireless sensor network increases and the data loss and the delay factors are reduced.*

**Keywords:** AODV Routing, RORCA, Recall, Precision, Throughput, Pegasus, Omnet++.

## I INTRODUCTION

The advancements in wireless communication technologies enabled large scale Wireless Sensor Networks (WSNs) deployment. Due to the feature of ease of deployment of sensor nodes, wireless sensor networks (WSNs) have a vast range of applications such as monitoring of environment and rescue missions. Wireless sensor network is composed of large number of sensor nodes. The event is sensed by the low power sensor node deployed in neighborhood and the sensed information is transmitted to a remote processing unit or base station.

Wireless sensor networks (WSN) are playing-vital role in commercial, military application, critical event monitoring and networks monitoring such as fire detection in forest, gas monitoring in coal mining, large scaled wireless sensor nodes are deployed in wide range of area. The sensor nodes collectively work together and send the detected information to other sensor nodes. To transmit the message it needs twice the energy it takes to receive the message. Major challenge of WSN is reducing the energy consumption. To increase the efficiency of energy using a techniques called inter-cluster communication and Rational Operating Recall Curve Algorithm (RORCA). In inter-cluster communication, each cluster has a coordinator referred to as a cluster head. In rational operating recall curve algorithm, the intrusion detection rate of the system can find effectively.

### 1.1 Clustering in WSN

Clustering mechanism helps in improving the energy efficiency in sensor networks. Clustering in WSN splits the sensor nodes into small groups. A cluster head (CH) would be selected for each group. The nodes in the group reports the information gathered to the CH. All sensor nodes in a cluster are called cluster members; including one cluster head. Cluster heads transmit the aggregated data to a base station, instead of collecting data from cluster members. Thus, clustering reduces the overhead of the sink.

Clustering is introduced in network and is divided into 'n' radial level with base station (BS) at the center

#### Cluster Head

Cluster

Sink

### Figure 1: Cluster Model

The main idea of clustering is optimal selection of cluster head (CH) based on their level. Then the node having more energy is selected as cluster head.

### 1.2 AODV Routing Protocol

Ad-hoc On-demand Distance Vector (AODV) routing protocol uses traditional routing information-base (RIB) for one task creation and to determine an up-to-date path to receiver.

The neighbors are notified in case of route is broken due to temporal packet loss. Messages can be controlled using header field named control and breakages of route are as follows:

- Route Request Message (RREQ)
- Route Reply Message (RREP)

- Route Error Message (RERR)
- HELLO Messages.

## II RELATED WORK

Lettieri and Srivastava [6] have proposed a new adaptive link layer control technique to provide robust and energy efficient operation even in the presence of orders of magnitude variations in bit error rates. An exchange exists between one another to reduce the header and physical layer overhead by making frames large. In addition, the adaptive frame length control can be used to improve the energy efficiency for a desired level of good put, and to extend the usable radio range with graceful throughput degradation, but has battery energy limitations.

Ferriere et al. [4] have proposed the Simple Packet Combining (SPaC) error-correction scheme for wireless sensor networks. When two or more corrupt versions of a packet grabs, a packet merge sequence makes an effort to achieve the original packet from the corrupt copies. Packet combining exploits a multi-hop wireless network and point-to-point forward error correction (FEC), packet combining therefore helps multi-node interactions such as multi-hop routing or broadcasting as well as to hop-by-hop communication. SPaC does not estimate the channel conditions.

Dong et al. [1] have proposed a Dynamic Packet Length Control (DPLC). DPLC is used for packet length optimizations for sensor networks to make more efficient in terms of channel utilization, incorporate a lightweight and accurate link estimation method. It also provides two easy use of services, one is aggregating a small message and another is fragmenting large message to facilitate upper layer application programming. A DPLC overcomes the limitations of prior work, but still can reduce the overheads and increase the energy efficiency by using some other techniques.

Krishnan et al. [5] have proposed a packet length adaptation (PLA) in wireless local area network (LAN). A packet can be lost due to many reasons, to reduce some of loss in medium access control (MAC) layer using a local packet adaptation algorithm. In PLA, each node dynamically adjusts its packet length based on estimates of the probabilities of each significant type of packet loss. In PLA, the access point periodically broadcasts channel information in order to estimate current network conditions. The throughput gains up to 20% via NS-2 simulations, but in transmitting hidden nodes unable to sense the transmission which leads to staggered collision.

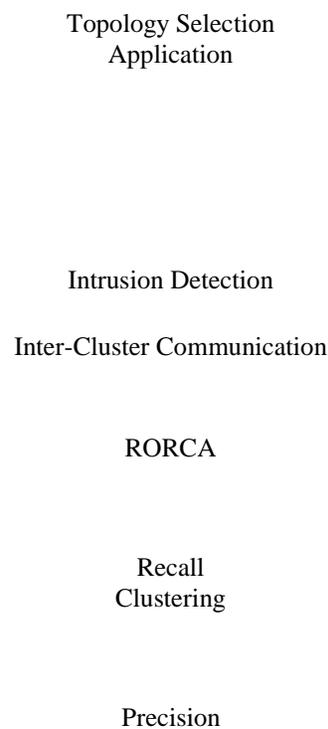
Jamieson and Balakrishnan [2] have proposed a partial packet recovery (PPR). In before, to correct small number of bit errors, retransmitting the whole packet by using Forward Error Correction (FEC). The FEC wasting network capacity, to overcome this inefficiency, jamieson and balakrishnan implement a partial packet recovery (PPR) system. PPR incorporate two new ideas such as SoftPHY and a post-amble scheme. As a result increases end-to-end capacity by a factor of  $2\times$  under moderate load, but still can improve the performance of routing protocol in SoftPHY.

## III DESIGN

It is proposed that the intrusion detection and fusion methods, to overcome the active and the passive attacks in the network. Rational Operating recall Curve Algorithm (RORCA) is introduced in the rational intrusion detection method. This Rational operating recall Curve Algorithm is used to find the intrusion detection rate of the system

effectively. And also employs to reduce the intrusion rate of the system during the process of communication in the network. The fusion method is also introduced to find the precision rate of the network. As a result, the proposed method performs well in monitoring the network and reduces the data loss.

Figure 2 shows an overall architecture design of proposal method. The application or data or messages are passed. Network is formed and route is discovered, then sending a request to topology. A topology is selected as random topology. In Inter-cluster communication cluster head is selected and cluster is formed. Then a method called Intrusion detection used for prevent from attackers. In intrusion detection Rational Operating Recall Curve Algorithm is used, then followed by recall, precision and fusion method.



**Figure 2: Architecture.**

#### **· Network Creation**

The network is created based on topology. Here using random topology i.e., mostly using Mesh topology. The mesh topology combines one or more topologies.

#### **· MAODV Protocol**

The Multi-casting AODV routing protocol builds on top of the DSDV protocol that was previously described. AODV is an improvement of DSDV as it minimizes the number of required broadcasts since it creates routes in an

on-demand basis, in contrast to DSDV which maintains a complete set of routes. It utilizes destination sequence numbers to ensure loop-freedom at all times and to avoid the count-to-infinity problem associated with classical distance-vector protocols.

When a node needs a route to a destination it broadcasts a Route Request (RREQ) message. The RREQ message is spread throughout the network and as soon as the message reaches a node with fresh enough routes to the specific destination or the destination node itself, a Route Reply (RREP) message is unicasted back to the requesting node.

AODV is able to provide unicast, multicast and broadcast communication ability. This capability of having all three communication forms in a single protocol offers numerous advantages. When searching by using the multicast route discovery it increases the unicast routing knowledge and vice versa. By having all three communication forms in a single protocol simplifies the implementation process of the protocol.

### · Energy Aware Clustering

PEGASIS (Power-Efficient Gathering in Sensor Information Systems) is for each node to receive from and transmit to close neighbors and take turns being the leader for transmission to the BS. This approach will distribute the energy load evenly among the sensor nodes in the network. We initially place the nodes randomly in the play field, and therefore, the  $i$ -th node is at a random location. The nodes will be organized to form a chain, which can either be accomplished by the sensor nodes themselves using a greedy algorithm starting from some node. Alternatively, the BS can compute this chain and broadcast it to all the sensor nodes. We used random 100-node networks for our simulations.

PEGASIS performs data fusion at every node except the end nodes in the chain. Each node will fuse its neighbor's data with its own to generate a single packet of the same length and then transmit that to its other neighbor. With our simulation experiments, found that the greedy chain construction performs well with different size networks and random node placements. In constructing the chain, it is possible that some nodes may have relatively distant neighbors along the chain. Such nodes will dissipate more energy in each round compared to other sensors. The performance of PEGASIS is improved by not allowing such nodes to become leaders. By setting a threshold on neighbor distance to be leaders, able to slightly improve PEGASIS's performance further by applying a threshold adaptive to the remaining energy levels in nodes. Whenever a node dies, the chain will be reconstructed and the threshold can be changed to determine which nodes can be leaders.

### 3.4 Intrusion Detection

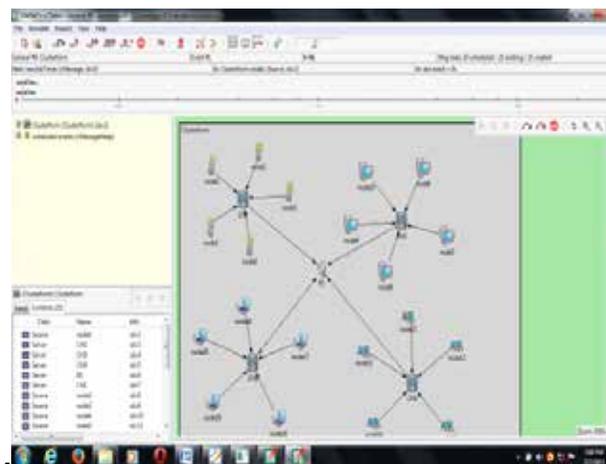
Intrusion Detection is used for detecting the intrusion, which are created by the external environment. To detect the intrusion in network proposing the Rational Operating recall Curve Algorithm (RORCA). This Rational operating recall Curve Algorithm is used to find the intrusion detection rate of the system effectively and also reduce the intrusion rate of the system during the process of communication in the network. The data fusion method is used to increase the precision rate of the system.

## IV SIMULATION

This project is simulated in OMNeT++. OMNeT++ is an object-oriented modular discrete event network frameworks simulation. OMNeT++ simulations can be run under various user interfaces. OMNeT++ also supports parallel distributed simulation. The parallel simulation algorithm can easily be extended, or new ones can be plugged in. OMNeT++ simulation tool is chosen, because it provides better performance than NS2 and OPNET. OMNeT++ is much scalable than NS2. OMNeT++ joins with MIXIM and used in numerous domains from queuing network simulations to wireless and ad-hoc network simulations. OMNeT++ cost is also low and it works well and coding of this also easy compared to test-bed like TOSSIM. OMNeT++ provides more accuracy of results, it's a new development tool and it becomes popular in future.

### 4.1 Simulation Results

The simulation results are shown in figure 3 and 4 respectively.



**Figure 3: Network Creation**

In fig. 3, the network is created by random topology selection which contains many numbers of cluster heads and up to 90's to 100's sensor nodes are connected.

In fig. 4, the cluster communication is done by using pegasis clustering. In this simulation result, there is transmission of message between cluster heads and sensor nodes.

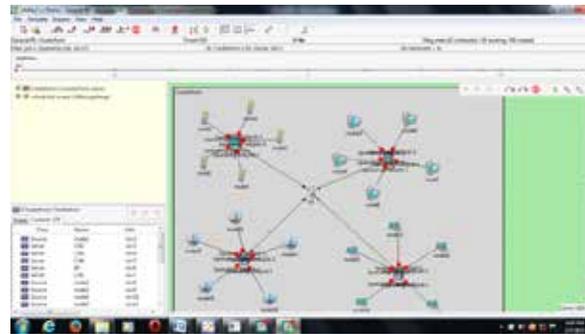
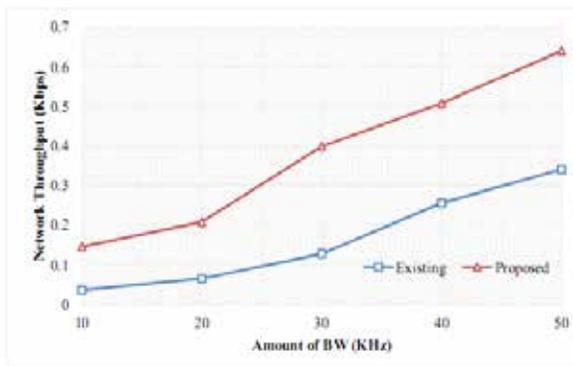


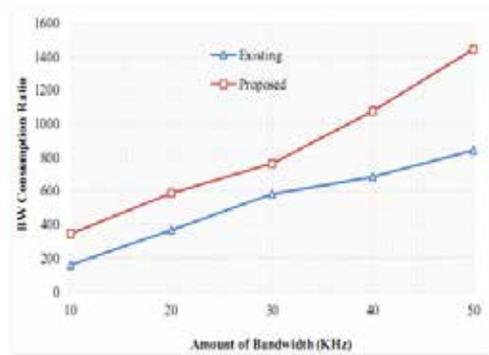
Figure 4: Clustering

4.2 Expected performance Analysis

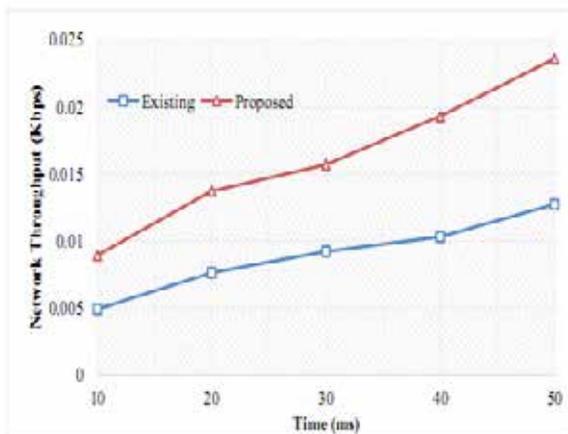
In this section, representing expected performance analysis in comparison with proposal and existing method.



(a) Network Throughput vs Bandwidth.



(b) Amount of Bandwidth vs BW Ratio.



(c) Network Throughput vs Time.

Figure 5: Performance Comparisons.

In fig. 5, has three performance simulations in comparison of existing and proposed method. (a) represents network throughput in kbps versus amount of bandwidth (BW) in kHz. (b) represents bandwidth consumption ratio versus amount of bandwidth in kHz. (c) represents network throughput in kbps versus time in milliseconds.

## V CONCLUSION

This paper presents PEGASIS Clustering and Rational Operating recall Curve Algorithm (RORCA). PEGASIS Clustering is used for improving performance and aware energy. RORCA is introduced in the rational intrusion detection method. This Rational operating recall Curve Algorithm is used to find the intrusion detection rate of the system effectively. The fusion method is also introduced to find the precision rate of the network. As the result, it is expected that the proposed method performs well when compared with the existing method. Also the energy efficiency increases by using pegasis clustering and provides safety by preventing from the attackers.

## REFERENCES

- [1] Bu, J, Dong, W, Liu, X, Chen, C, He, Y, Chen, G and Liu, Y, DPLC-Dynamic packet length control in Wireless sensor networks: IEEE INFOCOM, 2010, pp. 1-9.
- [2] Balakrishnan, H, and Jamieson, K, PPR- partial packet recovery for wireless sensor networks: ACM SIGCOMM, 2007, vol. 37, Issue. 4, pp. 1-5.
- [3] Davis, J & Goadrich, M, Relationship between precision-recall curves: 23<sup>rd</sup> IEEE conf. on machine learning, 2006, pp. 3-4.
- [4] Estrin, D, Ferriere, D and Vetterli, M, Packet combining in sensor networks: ACM SenSys,2005, pp. 102-115.
- [5] Haghani, E, Krishnan, N, and Zakhor, A, Packet length adaptation in WLANs with hidden nodes and time-varying channels: IEEE GlobeCom, 2011, pp. 1-6.
- [6] Lettieri, P and Srivastava, B, Adaptive frame length control for improving wireless link throughput, range, and energy efficiency: IEEE INFOCOM, 1998, vol. 2, pp. 21-25.

# OPTIMIZING COST USING KNIGHT TOUR FOR CONTENT ADAPTIVE STEGANOGRAPHY

S.Monisa<sup>1</sup>, P.Ganesh Kumar PhD<sup>2</sup>

<sup>1</sup>PG Scholar, <sup>2</sup>Assistant Professor, Dept of IT Anna University Coimbatore (India)

## ABSTRACT

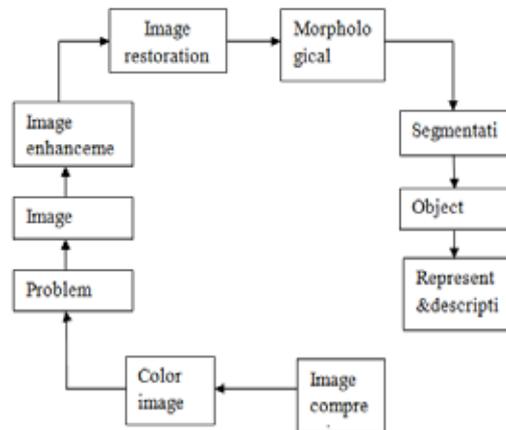
*Steganography includes all the actions that must be transported out to hide and protect the secret data inside the cover image encryption. In the first step of the embedding phase, the plain text will be encrypted, there are several encryption methods that can be applied to encrypt the data, but in this situation, we need a method that does not produce a cipher text longer than the plain text. The embedding algorithm is the most prominent part of the steganographic methods. In fact, it defines which pixels of the image should be changed and also in what order they will be altered with the secret data. The embedding algorithm of the proposed steganographic method is based on the described “Knight Tour” algorithm. The “Knight tour” algorithm is a suitable technique to formulate the sequence of the secret bit stream within the image pixels. When the sequence of the target pixels is defined in the previous step, now it’s the time to replace the least significant bits of the image pixels with the bit stream of the secret message.*

**Keywords :** *Steganography, Encryption, Knight Tour, PSRM, PPM,*

## I. INTRODUCTION

Steganography has been an important subject since people started communicating in writing. Steganography means hiding a secret message the embedded message within a larger one source cover in such a way that an observer cannot detect the presence of contents of the hidden message. Today the growth in the information technology, especially in computer networks such as Internet, Mobile communication and Digital Multimedia applications such as Digital camera, handset video etc. has opened new opportunities in scientific and commercial applications. But this progress has also led to many serious problems such as hacking, duplications and malevolent usage of digital information. Steganography finds its role in attempt to address these growing concerns.

The basic idea is to use the values of pixel pair as a reference coordinate, and search a coordinate in the Neighborhood set of this pixel pair according to a given message digit. The pixel pair is then replaced by the searched coordinate to conceal the digit. The main goal of steganography is to communicate securely in a completely undetectable manner and to avoid drawing suspicion to the transmission of a hidden data. It is not to keep others from knowing the hidden information, but it is to keep others from thinking that the information even exists. The basic model of steganography consists of Carrier, Message and Password. Carrier is also known as cover-object which the message is embedded and serves to hide the presence of the message.



**Fig 1.1 Key stages of Digital Image processing**

The key stages of image processing is image acquisition, image enhancement, image restoration, morphological processing, segmentation, object recognition, and represent and description can be performed in images. These stages can be processed from color images or any other images.

Relating the embedding cost in a distortion function to statistical detectability is an open vital problem in modern steganography we propose some rules for ranking the priority profile for spatial images. Following such rules, we propose a five-step cost assignment scheme. Previous steganographic schemes, such as HUGO, WOW, S-UNIWARD, and MG, can be integrated into our scheme.

## II. RELATED WORK

In an attempt to alleviate the negative impact of unavailable cover model, some steganographic schemes utilize the knowledge of the so-called “precover” when embedding secret data. The precover is typically a higher resolution (unquantized) representation of the cover, such as the raw sensor output before it is converted to an 8-bit per channel color image. The precover object is only available to the sender but not to the Warden, which seems to give a fundamental advantage to the sender. Provide theoretical insight for why side-informed embedding schemes for empirical covers might provide high level of security. By adopting a piece-wise polynomial model corrupted by AWGN for the content, we prove that when the cover is sufficiently non-stationary, embedding by minimizing distortion with respect to the precover is more secure than by preserving a model estimated from the cover (the so-called model-based steganography). Moreover, the side-informed embedding enjoys four times lower steganographic Fisher information than LSB matching.

A standard way to design steganalysis features for digital images is to choose a pixel predictor, use it to compute a noise residual, and then form joint statistics of neighboring residual samples (co-occurrence matrices). Proposes a general data-driven approach to optimizing predictors for steganalysis. First, a local pixel predictor is parameterized and then its parameters are determined by solving an optimization problem for a given sample of cover and stego images and a given cover source. Our research shows that predictors optimized to detect a specific case of steganography may be vastly different than predictors optimized for the cover source only. The results indicate that optimized predictors may improve steganalysis by a rather non-negligible margin.

Furthermore, we construct the predictors sequentially having optimized  $k$  predictors, design the  $k + 1$ st one with respect to the combined feature set built from all  $k$  predictors. In other words, given a feature space image model extend diversify the model in a selected direction functional form of the predictor in a way that maximally boosts detection accuracy.

Steganography has been an important subject since people started communicating in writing. Steganography means hiding a secret message the embedded message within a larger one source cover in such a way that an observer cannot detect the presence of contents of the hidden message. Today the growth in the information technology, especially in computer networks such as Internet, Mobile communication, and Digital Multimedia applications such as Digital camera, handset video etc. has opened new opportunities in scientific and commercial applications. But this progress has also led to many serious problems such as hacking, duplications and malevolent usage of digital information. Steganography finds its role in attempt to address these growing concerns. We know that, with the use of steganographic techniques, it is possible to hide information within digital images and video files which is perceptually and statistically undetectable. Proposes a new data-hiding method based on pixel pair matching (PPM). The basic idea of PPM is to use the values of pixel pair as a reference coordinate, and search a coordinate in the neighborhood set of this pixel pair according to a given message digit. The pixel pair is then replaced by the searched coordinate to conceal the digit.

The traditional way to represent digital images for feature based steganalysis is to compute a noise residual from the image using a pixel predictor and then form the feature as a sample joint probability distribution of neighboring quantized residual samples the so-called co-occurrence matrix. In this paper, we propose an alternative statistical representation instead of forming the co-occurrence matrix; we project neighboring residual samples onto a set of random vectors and take the first-order statistic histogram of the projections as the feature. When multiple residuals are used, this representation is called the projection spatial rich model (PSRM). On selected modern steganographic algorithms embedding in the spatial, JPEG, and side-informed JPEG domains, we demonstrate that the PSRM can achieve a more accurate detection as well as a substantially improved performance versus dimensionality trade-off than State-of-the-art feature sets. Enhancing the security of the traditional LSB matching, two improved LSB-matching methods are proposed. In the steganographical procedure, the Markov chain distance based on the second-order statistics is chosen as the security metric to control the modification directions of  $\pm 1$  embedding. The first method is based on stochastic modification, which directly determines the modification directions by the empirical Markov transition matrix of a cover image and the pseudorandom number generated by a pseudorandom number generator. The second one is based on genetic algorithm, which is used to find the optimum matching vector to make the security metric as small as possible. Experiments show the proposed algorithms outperform LSB matching and LSB replacement in a sense of the first order and second-order security metrics. And the adjacent calibrated COM-HCF steganalytical tests also show that the two algorithms are more secure than the traditional ones.

### III. SYSTEM ARCHITECTURE DESIGN

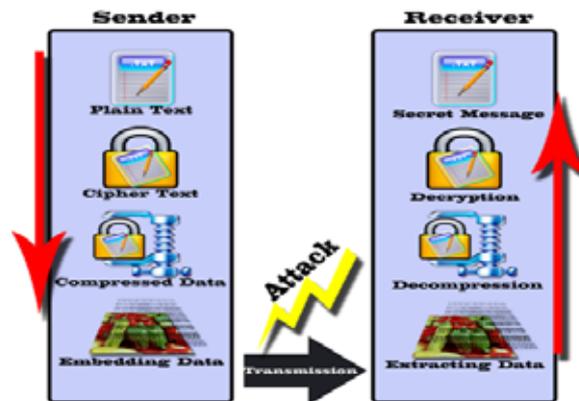


Fig 3.1. Architecture design

#### 3.1 Module Description

- Two-phase perspective on cost assignment and Different techniques of Steganography
- New cost assignment scheme
- Experimental results using Stego images

#### 3.2 Two-Phase Perspective On Cost Assignment

We explore a different perspective on the cost assignment scheme. We conceptually separate the assignment process into two phases. First, we sort image elements in an order where the elements after modification leading to better undetectability are ranked higher. We call the ranking order as priority profile.

In the second phase, we assign cost-values, which follow a specified distribution (called a cost - value distribution), to the corresponding image elements. In this way, the security impact of the cost assignment can be regarded as being determined by two factors — priority profile and cost-value distribution. If these two factors are independent, we may optimize them separately to enhance the security performance. In the following subsections, we will investigate their roles in steganographic security.

##### 3.2.1 Different techniques of Steganography

###### A. Spatial Domain based Steganography

It includes LSB (Least Significant Bit) Steganography. The spatial methods are most frequently employed because of fine concealment, great capability of hidden information and easy realization. LSB Steganography includes two schemes:

- Sequential Embedding

- Scattered Embedding.

### **B. Transform Domain based Steganography**

The method of transform domain Steganography is to embed secret data in the transform Coefficients.

### **C. Document based Steganography**

This method embeds data in documents files by adding tabs or spaces to .txt or .doc files.

### **D. File Structure based Steganography**

This method inserts secrets data in the redundant bits of cover files, such as the reserved bits in the file header or the marker segments in the file format.

## **3.3 NEW COST ASSIGNMENT SCHEME**

First present a new cost assignment scheme. Then explain the naming syntax for different options in the proposed scheme. Then discuss how to enhance the performance of previous methods by integrating them into our scheme. Finally presenting the general method to enrich the ways of obtaining priority profile.

## **3.4 Experimental Results Using Stego Images**

The experiments in this section are conducted on BOSS Base version 1.01 with 10000 gray scale images of size  $512 \times 512$ . The performances are evaluated by using steganalyzer with a 32671-D SRM feature set and an ensemble classifier, where Fisher linear discriminates is used as base learners. A number of 5000 randomly selected cover images and their Stego counterparts are used for training, while the rest 5000 cover images and their Stego counterparts are used for testing. The performance is evaluated by the testing error, which is the average of the false positive rate and the false negative rate, and we find that the testing error is usually slightly lower than the ensemble's "out-of-bag" error.

## **IV. IMPLEMENTATION**

The MATLAB language supports the vector and matrix operations that are fundamental to engineering and scientific problems. It enables fast development and execution. With the MATLAB language, you can program and develop algorithms faster than with traditional languages because you do not need to perform low-level administrative tasks, such as declaring variables, specifying data types, and allocating memory. In many cases, MATLAB eliminates the need for 'for' loops. As a result, one line of MATLAB code can often replace several lines of C or C++ code. At the same time, MATLAB provides all the features of a traditional programming

language, including arithmetic operators, flow control, data structures, data types, object-oriented programming (OOP), and debugging features.

MATLAB is a wide range of applications, including signal and image processing, communications, control design, test and measurement, financial modeling and analysis, and computational biology. Add-on toolboxes (collections of special-purpose MATLAB functions, available separately) extend the MATLAB environment to solve particular classes of problems in these application areas. MATLAB provides a number of features for documenting and sharing your work. You can integrate your MATLAB code with other languages and applications, and distribute your MATLAB algorithms and applications.

#### 4.1. Steganographic Algorithm

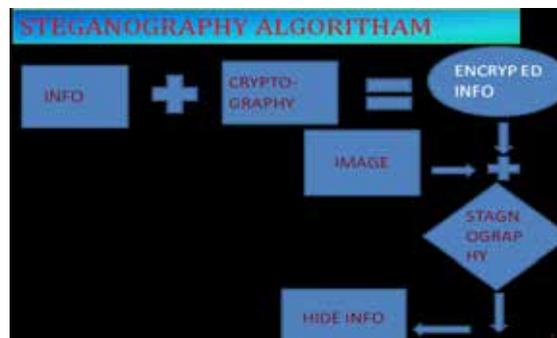


Fig 4.1 Steganography Algorithm

The information is combined with the image and it is encrypted by cryptography and then compressed these things and it changes into steno-graphy which hides the information.

#### 4.2 Knight Tour Algorithm

The “Knight tour” algorithm is a suitable technique to formulate the sequence of the secret bit stream within the image pixels. The advantage of the knight tour method over the PRNG technique is that, it is a self-developed algorithm based on the knight tour mathematical problem and it is almost unknown for the unintended receivers. By considering the image as an extended chessboard, we can have an algorithm, which determines the path of the knight within the image. The solution of the “Knight Tour” problem divides the chess board into the blocks with the size of 4x4 squares.

#### 4.3 Performance Analysis

The general architecture of the proposed steganography method is designed and implemented. The whole process is composed of two main phases, which are embedded phase and Receiving phase.

- **Embedding Process** - This phase includes all the activities that must be carried out to hide and protect the secret data inside the cover image. The sender uses some algorithms to encode and compress the data and then embeds the bit stream into the image. Moreover, the secret key is defined as the first position of the bit stream within the image. This key is identified just for the sender and receiver.
- **Receiving Phase** - On the other side of the communication line, the receiver should be able to comprehend the secret data within the Stego image. Therefore, another procedure is required to recover the content of the message and restructure it. First of all, based on the stego key and the extracting algorithm (the same as sender side) the bits of the secret message are obtained to compose a compressed data. Then the unzipping algorithm will generate the encrypted data and finally the plain message will be revealed.

## V. RESULT

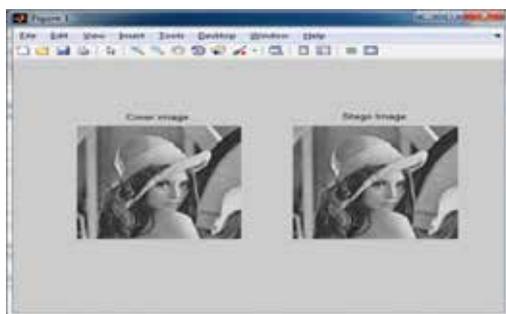


Fig 5.1 Encrypted Image

```
Command Window
Hiding message... Please wait.
Done.
SNR Signal to Noise Ratio
37.4684

PSNR Peak Signal to Noise Ratio
42.7777

Recovering hidden message... Please wait.
Done.
Bit error rate
0
fx >>
```

5.2 Hiding Images

```
Command Window
Enter Input Text = stegography
Enter Key Value 6
Encryption Result

ENCRYPT =

ydkunagvne

isOK =

1

ans =

kh
ho
oo
or
fx >>
```

Fig 5.3 Encrypted Data

## VI. CONCLUSION

In this work, we decided to fix the weakness of the Simple LSB system by providing some enhancements. The Enhanced LSB method utilizes three fundamental improvements specifically embedding algorithm, encryption and compression. The process starts with the encoding the confidential information by using encryption techniques. Both of the sender and receiver have a secret key which is used in encryption and decryption phases. Afterward, the compression technique reduces the size of encrypted data to improve the payload capacity. Clearly, as much the length of input data increases, the rate of compression surges, as well. Finally, the generated bit streams are embedded into the image in the positions which are defined by the proposed

embedding algorithm. The aforesaid embedding method is an extended form of proposed algorithm and provides the maximum number of pixels to hide the secret message.

## REFERENCES

- [1] Böhme R. (2010) *Advanced Statistical Steganalysis*. Berlin, Germany: Springer-Verlag.
- [2] Bin Li, Member, Shunquan Tan, Member, Ming Wang, and Jiwu Huang, Senior Member, (2014) 'Investigation on Cost Assignment in Spatial Image Steganography', *IEEE Transactions On Information Forensics And Security*, vol. 9, no. 8.
- [3] Chen L. Shi Y.Q Sutthiwan P and Niu X. (2013) 'A novel mapping scheme for steganalysis', in *Digital Forensics and Watermarking* (Lecture Notes in Computer Science), vol. 7809. Berlin, Germany: Springer-Verlag, pp. 19–33.
- [4] Chonev V.K and Ker A.D. (2011) 'Feature restoration and distortion metrics', *Proc. SPIE*, vol. 7880, pp. 0G01–0G14,
- [5] Fridrich J. and Kodovský J. (2012) 'Rich models for steganalysis of digital images', *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, and pp. 868–882.
- [6] Fridrich J. (2009) *'Steganography in Digital Media Principles', Algorithms, and Applications* Cambridge, U.K, Cambridge Univ. Press.
- [7] Holub V. Fridrich J. and Denmark T. (2013) 'Random projections of residuals as an alternative to co-occurrences in steganalysis', *Proc. SPIE*, vol. 8665, p. 86650L.
- [8] Huang F, Huang J. and Shi Y.Q. (2012) 'New channel selection rule for JPEG steganography', *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 4, pp. 1181–1191.
- [9] Ker A. D *et al.*, (2013) 'Moving steganography and steganalysis from the laboratory into the real world', in *Proc. 1st ACM Workshop Inform. Hiding Multimedia Security*, pp. 45–58.
- [10] Kodovský J. Fridrich J. and Holub V. (2012) 'Ensemble classifiers for steganalysis of digital media', *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 432–444.
- [11] Li B.J. He Huang J. and Shi Q.Y (2011) 'A survey on image steganography and steganalysis', *J. Inform. Hiding Multimedia Signal Process* vol. 2, no. 2 pp. 142–172.
- [12] Mielikainen J. (2006) 'LSB matching revisited', *IEEE Signal Process. Lett.* vol. 13, no. 5, pp. 285–287.
- [13] Pevný T. Bas P. and Fridrich J. (2010) 'Steganalysis by subtractive pixel adjacency matrix,' *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 2, pp. 215–224.