EMERGING TRENDS IN SCIENCE, ENGINEERING & TECHNOLOGY (ICETSET-2018)

Editors: Prof. (Dr.) Anuranjan Misra, Ms. Pushpa & Ms. Shahina Anjum

Organised By:

ACCURATE INSTITUTE OF MANAGEMENT & TECHNOLOGY

49, Knowledge Park – III, G.Noida, Uttar Pradesh, INDIA
Web: www.accurate.in, Toll Free: 1800-180-3515, Ph. 0120-2328234/35

International VSRD
J O U R N A L S

VSRD INTERNATIONAL JOURNAL OF TECHNICAL AND NON-TECHNICAL RESEARCH

INTERNATIONAL CONFERENCE

ON

"EMERGING TRENDS IN SCIENCE, ENGINEERING & TECHNOLOGY"

(ICETSET-2018)

EDITORS

Prof. (Dr.) Anuranjan Misra
Ms. Pushpa
Ms. Shahina Anjum

ORGANISED BY

ACCURATE INSTITUTE OF MANAGEMENT & TECHNOLOGY

INTERNATIONAL CONFERENCE ON "EMERGING TRENDS IN SCIENCE, ENGINEERING & TECHNOLOGY" (ICETSET-2018)

Printed & Bound in India

E D I T O R I A L

"International Conference on Emerging Trends in Science Engineering & Technology (ICETSET-2018)" *is an all-encompassing term with comprehensive connotations for Technology & Innovation, Big Data, Cloud Computing, Internet of Things, Information Security, Electronic CAD Tools & Auto CAD, Smart Vehicles & Smart Cities, VLSI Technology Trends, and many others. It requires various sectors to work in synchronization and complements each other.*

Accurate Institute of Management & Technology, Greater Noida, Uttar Pradesh, India is organizing two days International Conference on Emerging Trends in Science Engineering & Technology from 23-24 March 2018.

This conference will offer an excellent International platform for the academicians, researchers, engineers, industrial participants and budding students around the world to share and exchange concept of current developments, innovations and expansions on Technologies. In this conference, apart from invited talks, there are scopes for contributory talks and posters. Researchers from various disciplines are also welcomed for sharing their research findings in the conference.

We extend our heartiest gratitude to Accurate Institute of Management & Technology, Greater Noida for funding the conference and administrators, faculty members from different departments, research scholars, students and staff members for their precious co-operation throughout the conduct of this successful and prestigious event.

We wish all the global delegates, academicians, industrialists, researchers of the conference enjoying the academic program with fruitful interaction with the ambience of Accurate family and to have a pleasant stay.

Prof. (Dr.) Anuranjan Misra
(Director-AIMT & Conference Director)

T H E M E S

E-governance and government

Big data

Cloud and Green Computing

Data Mining

Computer Networking

Cyber Security

WSN

Green Computing & Internet of Things

Electrical Machines

Electromagnetic Compatibility

High Voltage and Insulation Technology

Advanced Power Semiconductor Devices

Antennas and Wave Propagation

Embedded System

Image Processing & Understanding

Optical Communications

Cognitive System Research

Radar and Microwaves

Network Control (Admission/Flow/Congestion Control, etc.)

Wired/Wireless Integrated Networks

Town Planning

Bridge Engineering and Building Structure

Materials and Construction Systems

Building Technology

Construction and Control; Detection and Transformation

Civil and Structural Engineering

Virtual Manufacturing

Nano Machines

Safety and Security in Automobile

Advanced and Digital Manufacturing

Dynamics and Vibration

High Power/Energy Batteries, Battery/Energy Management and Charging Systems

Electronics for Electrical Machine/Hybrid Engine Control Systems (hard and soft)

"Zero Emission" Range Extension Solutions

Quantum Computing

IT & Security Considerations

Sustainable Computing

Message from Chairman

I am delighted to note that Accurate Institute of Management & Technology has taken the initiative to provide an International platform to academicians, researchers, scientists and industrialists to share and discuss numerous themes under **"Emerging Trends in Science Engineering & Technology"**. *The theme of the conference being contemporary would evoke many views and perspectives and deliberations should prove to be thought provoking exercise.*

I convey my best wishes for the success of the conference.

I congratulate to AIMT Family for this endeavor.

With best regards!

C.L. SHARMA
Chairman
Accurate Group of Institutions

Message from Group Director

I am glad to know that an International Conference on **"Emerging Trends in Science Engineering & Technology"** *is being organized on March 23-24, 2018. This will be a significant day not only for Accurate Institute of Management & Technology but also for everyone associated with it. Events like this involve the meeting of best minds, entrepreneurs, leading to meaningful debates and fruitful interactions.*

Conferences, like these are not only the platform for exchanging of ideas but also are a 'Samagam' of researchers, scientists and industrialists.

I wish all the success to this conference.

Kind regards!

Ms. POONAM SHARMA
Group Director
Accurate Group of Institution

Message from Director

*I welcome all the global delegates to International Conference on **"Emerging Trends in Science Engineering & Technology"** being organized on March 23-24, 2018 in Accurate Institute of Management & Technology, Greater Noida. This conference will provide a platform for the research communities in the field of computer's trends & development from different areas. This conference is aimed at communicating current progress in the fields of Science, Technology & Innovation, Cloud Computing, Internet of Things, Information security, Electronic CAD tools & Auto CAD, Smart Vehicles & Smart Cities, VLSI Technology Trends, Configurable Computing, Optical Computing, Molecular Nanotechnology and many more. This ongoing International Conference is truly meant for scientific exchange and generation of new ideas in the chosen field along with personal interaction and social get together. This conference also aims at building bridges to strengthen global collaborations in many recent trends that are global in nature.*

I offer my warm welcome to all invited delegates to Accurate Institute of Management & Technology, Greater Noida, a splendid temple of learning.

Prof. (Dr.) ANURANJAN MISRA
Director
Accurate Institute of Management & Technology

Message from Conference Director Planning

I am happy to see that Accurate Institute of Management & Technology, Greater Noida is organizing an International Conference on **"Emerging Trends in Science Engineering & Technology"** from March 23 to March 24, 2018, which is a significant step for imparting technical advancement to its participants.

This conference brings-in eminent speakers and researchers from all walks of life together to exchange the knowledge. This conference will cover all relevant fields of Emerging Trends & Technologies with special emphasis on VLSI Technology Trends, Configurable Computing, Optical Computing, Molecular Nanotechnology, Cloud Computing and Big Data.

On behalf of AIMT family, I welcome all speakers and delegates to AIMT campus.

I wish this conference a grand success.

Prof. (Dr.) S.K. DUBEY
Director Planning
Accurate Institute of Management & Technology

ICETSET - 2018

ORGANIZING COMMITTEE

CHIEF PATRON

Mr. C.L. Sharma, Hon'ble Chairman
Ms. Poonam Sharma, Hon'ble Group Director

CONFERENCE DIRECTOR(S)

Prof. (Dr.) Anuranjan Misra, Director
Dr. S. K. Dubey, Director Planning

CONFERENCE CONVENER

Dr. A. K. Srivastava, Dean Academics
Dr. P. K. Talan, Dean Student Welfare
Mr. Arvind Kumar, Head ECE

CONFERENCE EDITORS

Dr. Anuranjan Misra
Ms. Pushpa
Ms shahina anjum

EXECUTIVE ORGANISATION COMMITTEE

Er Anil Ji Garg - Chairman, CSI Ghaziabad Chapter
Mr Saurabh Agrawal - Patron, CSI & Delivery Head, TCS
Prof Manoj K Pandey - Vice Chairman, CSI Ghaziabad Chapter
Dr Kavita Saxena - Hony Secretary, CSI Ghaziabad Chapter
Prof Ravi Singhal - Member, Managing Committee, CSI Ghaziabad Chapter
Dr. Shanker Lal Rajput, HOD, Applied Science, AIMT
Mr. Sunil Sharma, HOD, Computer Science & Engineering & IT, AIMT, Greater Noida
Mr. Amar Yadav, HOD, Mechanical Engineering, AIMT, Greater Noida
Mr. V. K. Pallaw, HOD, Computer Applications, AIMT, Greater Noida
Mr. Arvind Kumar, HOD, Electronics & Communication, AIMT, Greater Noida
Mr. Pranay Bhadauiya, HOD, Electrical & Electronics AIMT, Greater Noida

LOCAL ORGANISATION COMMITTEE

Mr. Mukesh Chauhan	*Mr. Manoj Bhatnagar*	*Dr. Swati Pawar*	*Mr. Sachin Dubey*	*Mr. Vivek Singh*
Mr. P. M. Tripathi	*Mr. Kumar Gaurav*	*Mr. Mohit Agarwal*	*Dr. Kapil Sharma*	*Ms. Seema Yadav*
Mr. Anand Singh	*Mr. Jagmohan Singh*	*Mr. Jitendra Singh*	*Mr. Adil Arfin*	*Mr. Anand Singh*
Dr. Raja Agarwal	*Mr. Hariom Tripathi*	*Mr. Lovelesh singh*	*Ms. Shailja*	*Mr. Awdhesh Jha*
Dr. Sadanand Singh	*Mr. Manish Sharma*	*Mr. Mitramani*	*Mr. Navin Jha*	*Mr. Sunil Yadav*
Mr. Yogesh Kaushik	*Ms. Ananta Srivastava*	*Dr. Sadanand Sing*	*Mr. Nikhil Mishra*	*Mr. Aditya Sharma*
Ms. Shivani Agarwal	*Mr. Rishikesh Kumar*	*Mr. H. P. Singh*	*Mr. Manish Giri*	*Mr. P. M. Tripathi*
Ms. Shivani Agarwal	*Mr. Gurpreet Singh*	*Mr. Ajit Yadav*	*Mr. Jayesh*	*Ms. Pratibha Singh*
Mr. A. M. Tripathi	*Ms. Poornima Sinha*	*Mr. S. D. Bhati*	*Mr. Vikrant*	*Ms. Pooja Sharma*
Mr. Manish Maurya	*Ms. Archana das*	*Mr. Ajit Yadav*	*Mr. Vikash Gautam*	*Mr. Manjeet Singh*
Mr. Prabhat Tiwari	*Mr. Deenanath Gupta*	*Mr. Amit Kumar*		*Ms. Vartika Rao*

C O N T E N T S

THE BRIDGE CLUSTERING ALGORITHM

¹Kamalpreet Bindra* and ²Dr. Anuranjan Mishra
¹Research Scholar, Department of Computer Science & Engineering,
Noida International University, G.Noida, Uttar Pradesh, INDIA.
²Prof. & Director, Accurate Institute of Management & Technology, G.Noida, Uttar Pradesh, INDIA.
*Corresponding Author: Kamalpreet.bindra@gmail.com

ABSTRACT

Clustering in data mining is a supreme step towards organizing data into some meaningful patterns. It plays an extremely crucial role in the entire KDD process also as categorizing data is one of the most rudimentary steps in knowledge discovery. Clustering is used for creating partitions or clusters of similar objects. Various clustering algorithms have been developed under different paradigms for grouping scattered data points and forming efficient cluster shapes with minimal outliers. In This paper a hybrid algorithm based on k-medoids and birch is proposed and termed as the bridge algorithm or TBA. The hybrid algorithm aims towards developing high quality clusters with accurate clustering results out of high dimensional data sets.

Keywords: Clustering, KDD, Outliers, Hybrid Algorithms, High Dimensional Data Sets.

1.   INTRODUCTION

Data is the goldmine in today's ever competitive world. Everyday large amount of information is encountered by organizations and people. An indispensible means to handle this data is to categorize or classify them into a set of groups, partitions or clusters. "Basically classification systems are either supervised or unsupervised ,depending on whether they assign new inputs to one of the finite number of discrete supervised classes or unsupervised categories respectively"[3][5]. Farley and Raftery (1998) suggest dividing the clustering algorithms in mainly two groups: hierarchical and partitioning methods, further HAN and KAMBER divide clustering techniques in : grid based methods and model based . The paper proposes to introduce a scalable categorical bridge algorithm which bridges the hierarchical and partitional approach towards clustering, for producing clustering of different sizes and shapes and reasonable removal of noise and outliers. The algorithm has the advantage that it can produce clustering's of different sizes in single execution and enables backtracking which is largely absent in many hierarchical algorithms.

A. Important Issues in Data Clustering

Clustering is a challenging field of research in which its potential applications pose their own special requirements. The following are typical requirements of clustering in data mining. Discovery of clusters with arbitrary sizes, shapes and densities: Many algorithms tend to find spherical clusters with similar sizes. They do not work well when clusters have different sizes. Clusters that have widely varying densities are harder to detect. It is important to develop algorithms that can detect clusters with arbitrary sizes, shapes and densities. As stated by [8] and [9].

Ability to deal with different types of attributes: The ability to analyse single as well as mixture of attribute types is demanded by real life applications.

Ability to cluster huge volume of data (scalability): Algorithms used to cluster huge volume of data should have linear or near-linear time and space complexities. Furthermore, algorithms that assume that all the data will fit in main memory are infeasible for large datasets.

Ability to cluster high dimensional data: Many clustering algorithms are good at having low dimensional data involving less than ten dimensions.

It is a challenge to cluster high dimensional data, especially considering that such data can be sparse and highly skewed.

Ability to deal with noise and outliers: Clustering algorithms should be able to handle outliers in order to improve cluster quality.

Finding subspace clusters: Clusters may exist in a subset of attributes. It is not feasible to simply look for clusters in all possible subsets of attributes for datasets having large number of attributes.

Order dependence: When the same dataset is presented in different order, the results produced by some clustering algorithms may become drastically different. While it would seem desirable to avoid such algorithms, sometimes the order dependence is relatively minor or the algorithm may have other desirable characteristics.

Problem Definition and Stages: Most existing clustering

methods use agglomeration methods to find out correlation clusters ,As stated by anil K Jain and Richard C Dubes. Algorithms for Clustering in Data. Prentice Hall, 1988[1] and Hierarchical Density-Based Clustering of Categorical Data and a Simplification billl Andreo 'oulos, Aijun An, and Xiaogang [4].

Clustering cannot be a one-step process. And is divided in the following stages:

- *Data Collection:* Includes the careful extraction of relevant data objects from the underlying data sources. In our context, data objects are distinguished by their individual values for a set of attributes.
- *Initial Screening:* Refers to the massaging of data after its extraction from the source or sources. This stage is closely connected to a process widely used in Data Warehousing, called Data Cleaning [4].
- *Representation:* Includes the proper preparation of the data in order to become suitable for the clustering algorithm. Here, the similarity measure is chosen, the characteristics and dimensionality of the data are examined.
- *Clustering Tendency:* Checks whether the data in hand has a natural tendency to cluster or not. This stage is often ignored, especially in the presence of large data sets.
- *Clustering Strategy:* Involves the careful choice of clustering algorithm and initial parameters, if any.
- *Validation:* Validation is often based on manual examination and visual techniques. However, as the amount of data and its dimensionality grow, we may have no means to compare the results with preconceived ideas or other clustering.
- *Interpretation:* This stage includes the combination of clustering results with other studies, e.g. classification, in order to draw conclusions and suggest further analysis.[6]

This list of stages is given for exposition purposes since I do not propose solutions for each one of them. Research mainly address the problem of Clustering Strategy by proposing a new and scalable algorithm for categorical data, and the problem of Clustering Tendency by proposing a heuristic for identifying appropriate values for the number of clusters that exist in a data set. Although the paper does not introduce new Validation techniques, I use a large number of measures, already given in clustering research, in order to achieve better results.

2.   PREVIOUS WORK

CLARANS (Clustering Large Applications based on Random Search) have been developed by Ng and Han (1994). This method identifies candidate cluster centroids by using repeated random samples of the original data. Because of the use of random sampling, the time complexity is O(n) for a pattern set of *n* elements. Combining partitional and hierarchical algorithms for robust and efficient data clustering with cohesion self merging[11].

In this paper, the authors discussed cohesion based clustering. The technique also employs hybrid approach and tolerance towards outliers in various fields. It has shown to be able to cluster data sets of arbitrary shapes. Hierarchical clustering algorithms for document datasets [12]

In this paper it is experimentally evaluated that nine agglomerative algorithms and six partitional algorithms are used to obtain hierarchical clustering solutions for document datasets. The research shows experimental results that partitional methods produce better hierarchical solutions than agglomerative methods. Combining with previous study on the effectiveness of the various partitional algorithms to produce k-way clustering solutions (Zhao and Karypis, 2004), it is showed that partitional methods are suitable for producing both flat and hierarchical clustering solutions for document datasets effectively and efficiently. A new class of agglomerative algorithms is also introduced by constraining the agglomeration process using clusters obtained by partitional algorithms. Experimental results showed that the constrained agglomerative methods improve the clustering solutions obtained by agglomerative or partitional methods alone. It is analysed that in most cases enforcing partitional cluster constraints improves the quality of the neighbourhood of each document, especially when the document has low similarities to others or has many documents with similar similarities. These improvements of neighbourhoods correlate well with the improvements of overall clustering solutions, which suggest that constrained agglomerative schemes benefit from starting with purer neighbourhoods.

Algorithms for clustering in data [anil k jain and Richard c dubesprentice hall] [13] In this paper authors contributions are two-fold in this paper. First, it discusses various approaches for measuring the dissimilarities/distances between interval data, investigate the relations among them, and present a comprehensive experimental study on clustering interval data. It show that the extended interval data clustering achieves better performance than traditional ones and produces more meaningful and explanatory results. Second, it proposes a two-stage approach for clustering interval data by exploiting the relations between the traditional distances and the modified distances. Experimental results show the effectiveness of their approach

Predicting customer purchase in an online retail business, a Data Mining Approach [14]

In this research, Aniruddha Mazumdar, May 2010, studied, implemented and analyzed some data mining tools as well as techniques and later analyzed the sampled data for interpretation. In this study data mining algorithms were used based on the clustering algorithms in conjunction with an

'Apriori' based Association rule mining algorithm. The research discusses different approaches that were used in

interpreting the results. The results clearly prove that:

- Using the VQ approach one can easily segment groups of Buyers based on the RFM values or all of them all together. However, this process needs the initials vectors as its input for it to form clusters.
- When predicting, different segmentations from the clusters formed can be used and the most populous ones are specifically focused on.
- Basing on the association rules alongside sufficient coverage, the product which the customer wishes to buy is predictable along with the purchase of particular products.

3. PHASES OF OPERATION

TBA has 2 phases of operations.

Phase -1: Defining the centroid, radius and diameter of a cluster. Given N d-dimensional data points in a cluster $\{X_i\}$ where i =1,2...N.the centroid $X0$,radius R and diameter D of the cluster are defined as :

$$\vec{X0} = \frac{\sum_{i=1}^{N} \vec{X_i}}{N}$$

$$R = \left(\frac{\sum_{i=1}^{N} (\vec{X_i} - \vec{X0})^2}{N}\right)^{\frac{1}{2}}$$

$$D = \left(\frac{\sum_{i=1}^{N} \sum_{j=1}^{N} (\vec{X_i} - \vec{X_j})^2}{N(N-1)}\right)^{\frac{1}{2}}$$

Here R, which is the radius is the average distance from data point to the centroid . D is the average pairwise distance within a cluster.with in the centroids of two clusters X0 and X1, the centroid eculidean distance D0 and manhattan distance D1 of the two clusters are defines as:

$$D0 = ((\vec{X0}_1 - \vec{X0}_2)^2)^{\frac{1}{2}}$$

$$D1 = |\vec{X0}_1 - \vec{X0}_2| = \sum_{i=1}^{d} |\vec{X0}_1^{(i)} - \vec{X0}_2^{(i)}|$$

Given N_1 d dimensional data points in a cluster : $\{X_i\}$ where i= 1,2 ,....N_1 and N_2 data points in another cluster $\{X_j\}$ where j= N_1 + 1, N_1 +2....N_1+N_2,the average cluster inter-cluster distance D2,average intra cluster distance D3 and variance increase distance D4 of any two clusters are defined as:

$$D2 = \left(\frac{\sum_{i=1}^{N_1} \sum_{j=N_1+1}^{N_1+N_2} (\vec{X_i} - \vec{X_j})^2}{N_1 N_2}\right)^{\frac{1}{2}}$$

$$D3 = \left(\frac{\sum_{i=1}^{N_1+N_2} \sum_{j=1}^{N_1+N_2} (\vec{X_i} - \vec{X_j})^2}{(N_1+N_2)(N_1+N_2-1)}\right)^{\frac{1}{2}}$$

$$D4 = \sum_{k=1}^{N_1+N_2} (\vec{X_k} - \frac{\sum_{l=1}^{N_1+N_2} \vec{X_l}}{N_1+N_2})^2$$

$$- \sum_{i=1}^{N_1} (\vec{X_i} - \frac{\sum_{l=1}^{N_1} \vec{X_l}}{N_1})^2 - \sum_{j=N_1+1}^{N_1+N_2} (\vec{X_j} - \frac{\sum_{l=N_1+1}^{N_1+N_2} \vec{X_l}}{N_2})^2$$

D3 is the D of merged clusters.X0,R,D are treated as properties of a single cluster and D0,D1,D2,D3,D4 are the properties between two clusters and calculated separately

Phase -2: Algorithm 1.1 (inserting data points into tree)initial tree T = 0, initial cluster iCi =0, where i is anything from 1 to N. The set iC contains data points from the data set. A set of nodes is are adjusted into root when root is full next data point becomes left child and hence forth right child. The left $node_i$ and right $node_j$.

1. begin
2. Initialize T=0 (current number of nodes), x,iC={}, Sc ={},C= {},R,D.$left_i$ = 0 , $right_j$ = 0. iC= {},Sc = {},N=0
3. for i= 1 to N
   For j = 1 to N
   Call iCi(i,j) ← construct left and right nodes
4. Sc(i,j) ← one level completed
5. Sv(i,j) ← update.
6. End

Algorithm 1.2 (cluster centroid , using k means)

1. Begin
2. Let X = $\{x_1,x_2,x_3,........,x_n\}$ be the set of data points and V = $\{v_1,v_2,.......,v_c\}$ be the set of centres.
1) Randomly select *'c'* cluster centres.
2) Calculate the distance between each data point and cluster centres
3) Assign the data point to the cluster centre whose distance from the cluster centre is minimum of all the cluster centres..
4) Recalculate the new cluster centre using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

where, *'c_i'* represents the number of data points in i^{th} cluster.

5) Recalculate the distance between each data point and new obtained cluster centre.
6) If no data point was reassigned then stop, otherwise repeat from step 3.

An important contribution is formulation of the clustering, problem in a way that, is appropriate for very large datasets by making the time and memory constraints explicit. In addition, TBA has the following advantages over previous distance-based approaches.

1. TBA is local (as opposed to global) in that each clustering decision is made without scanning all data points or all currently existing clusters .It uses measurements that reflect the natural closeness of points, and at the same time, can be incrementally maintained during the clustering process.
2. BIRCH exploits the observation that the data space is usually not uniformly occupied, and hence not every data point is equally important for clustering purposes.
3. A dense region of points is treated collectively as

a single cluster. Points in sparse regions are treated as outliers and removed optionally. BIRCH makes full use of available memory to derive the finest possible sub clusters (to ensure accuracy)while minimizing 1/0 costs (to ensure efficiency).The clustering and reducing process is organized and characterized by the use of an in-memory, height balanced and highly-occupied tree structure. Due to these features, its running time is linearly scalable. If we omit the optional Phase 4 5, TBA is an incremental method that does not require the whole dataset in advance, and only scans the dataset once.

4. CONCLUSION AND FUTURE SCOPE

The TBA algorithm is The bridge algorithm combines the features of BIRCH clustering algorithms whose feature of insertion and splitting is same as B-Tree algorithm and Partitioning clustering algorithm K-Medoids or PAM algorithm. The algorithm is applied on a typical super market dataset which is collected from a local supermarket. The TBA algorithm first make call to tree algorithm which is named as BCA algorithm that build a tree containing more than 1500 clusters on retail dataset. The insertion and splitting of this tree algorithm is same as B Tree algorithm but in this algorithm each node of the tree stores the node or tree label, the cluster number and the number of instances in that cluster. These large numbers of clusters are difficult to predict and understand. After that the algorithm make call to K-Medoids clustering algorithm which clusters the leaf nodes of the BCA clustering algorithm. Further experimentation and refinement is to be done on the algorithm to produce better clustering results.

5. REFERENCES

[1] Amandeep Kaur Mann, and Navneet Kaur, "Survey Paper on Clustering Techniques", IJSETR: International Journal of Science, Engineering and Technology Research (ISSN: 2278-7798), vol. 2, Issue 4, April 2013.

[2] B. Everitt, S. Landau, and M. Leese, *Cluster Analysis.* London: Arnold, 2001.

[3] RuiXu, Donald Wunch, survey of Clustering Algorithms", IEEE Transactions on Neural Networks, vol 16 no.3 may 2005.

[4] P. Berkhin. (2001) Survey of clustering data mining techniques. [Online].Available:http://www.accrue.com/products/rp_cluster_review.phthttp://citeseer.nj.nec.com/berkhin02survey.html

[5] J. Kleinberg, "An impossibility theorem for clustering," in *Proc. 2002Conf. Advances in Neural Information Processing Systems*, vol. 15, 2002, pp. 463–470.

[6] P. Sneath, "The application of computers to taxonomy," *J. Gen. Microbiol.*, vol. 17, pp. 201–226, 1957.

[7] D. Fasulo, "An analysis of recent work on clustering algorithms," Dept. Of Comput. Sci. Eng., Univ. Washington, Seattle, WA, Tech. Rep. 01-03-02, 1999.

[8] Tian Zhang, Raghu Ramakrishnan , MironLivny, BIRCH: an efficient data clustering method for very large databases, Proceedings of the 1996 ACM SIGMOD international conference on Management of data, p.103-114, June 04-06, 1996, Montreal, Quebec, Canada

[9] S. Guha, R. Rastogi, and K. Shim, "CURE: An efficient clustering algorithm for large databases," in *Proc. ACM SIGMOD Int. Conf. Management of Data*, 1998, pp. 73–84

[10] CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling , George Karypis , Eui-Hong Han, Vipin Kumar IEEE Computer 32(8): 68-75, 1999

[11] Ji He,Man Lan, Chew-Lim Tan, Sam-Yuan Sung, Hwee-Boon Low, "Initialization of Cluster refinement algorithms: a review and comparative study", Proceeding of International Joint Conference on Neural Networks [C]. Budapest, 2004

[12] Biswas, G., Weingberg, J. and Fisher, D.H., ITERATE: A conceptual clustering algorithm for data mining. IEEE Transactions on Systems, Man, and Cybernetics. v28C. 219-230

[13] Jiawei Han and Michheline Kamber, Data mining concepts and techniques-a reffrence book ,pg. no.-383-422.

[14] Arun K. Pujari,Data mining techniques-a reffrence book ,pg. no.-114-147.

[15] He, Z., Xu, X. and Deng, S., Scalable algorithms for clustering large datasets with mixed type attributes. International Journal of Intelligence Systems. v20. 1077-1089

[16] G. Ball and D. Hall, "A clustering technique for summarizing multivariatedata," *Behav. Sci.*, vol. 12, pp. 153–155, 1967

[17] M. Ester, H. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc.2nd Int. Conf. Knowledge Discovery and Data Mining (KDD'96)*, 1996,pp. 226–231.

[18] V. Cherkassky and F. Mulier, *Learning From Data: Concepts, Theory, and Methods.* New York: Wiley, 1998.

[19] A. Hinneburg and D. A. Keim. A general approach to clustering in large databases with noise. Knowledge and Information Systems 5(4):387-415,2003

[20] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[21] D. Cutting, J. Pedersen, D. Karger, and J. Tukey. Scatter/gather: A cluster-based approach to browsing large document collections. In Proceedings of the ACM SIGIR, pages 318–329, Copenhagen, 1992.

[22] S. Guha, R. Rastogi, and K. Shim. ROCK: a robust clustering algorithm for categorical attributes. I18n Proc. of the 15th Int'l Conf. on Data Eng., 1999.

[23] Ying zhao,George karypis,usama fayyad *Hierarchical clustering algorithms for document datasets*, min knowl disc(2005) 10 : 141.doi:10.1007/s10618-005-0361-3

[24] Abdellah Idrissi,Hajar Rehioui An improvement of denclue algorithm for the data clustering. Information & Communication Technology and Accessibility (ICTA), 2015 5th International Conference. IEEE Xplore 10 march 2016.

❏❏❏

CREATING STANDARD REPRESENTATION OF RTI RULES 2012 USING XML TECHNOLOGIES FOR ELECTRONIC INTERCHANGE OF DATA FOR INTERPRETATION BY SOFTWARE SYSTEMS

[1]Manoj Kumar Singh* and [2]Dr. Anuranjan Mishra
[1]Research Scholar, Department of Computer Science & Engineering,
Noida International University, G.Noida, Uttar Pradesh, INDIA.
[2]Prof. & Director, Accurate Institute of Management and Technology, G.Noida, Uttar Pradesh, INDIA.
*Corresponding Author Email ID: singhmk.in@gmail.com

ABSTRACT

In continuation of ongoing research on Electronic Interchange of legal data by creating standard representation of law & using big data set analysis for implementation in special reference to Right to Information Act 2005, a small study on creation of standard representation and test its effectiveness has been undertaken in context of India Law and create a framework for representation of law in order to achieve automated decision support systems and electronic exchange of data. The study is done by taking into consideration the Right to information rules 2012 which is derived from the Right to Information Act 2005. The RTI rule of encoded for standard representation using XML technologies. The XML schema is readable by software system for presentation, interpretation, validation etc. The modules of RTI rules that controls the first and second appeal has been encoded separately to test implementation. The objective of this study was to achieve standard representation of RTI rules 2012 a smaller segment related to RTI act but has multifold impact of justice delivery to public enshrined under RTI act 2005. The standard representation will be used test the encoding of process module of RTI the RTI application, 1st appeal and second appeal as well as replies provided by various authorities at different stages of justice delivery guaranteed under RTI which is providing information to public.

Keywords : Schema of law, India Legal System, RTI Rules 2012, First Appeal, XML, XSD, XSL Style sheet, HTML, Parsing, electronic interchange, Right to information, ICT.

1. INTRODUCTION

The proliferation of electronic commerce business model and technologies encouraged their application to the activities of government. The recent development of the digital India initiative of Government of India to integrate the government departments and the people of India spurred many developments in the field of IT and use of ICT technologies in government functions. Under Digital India programmes all ministries and departments shall offer their own services to the public Healthcare, Education, Judicial services etc. The major efforts have been taken in area of e-Governance, M-Governance, Development of IT infrastructure and communication channels.

Computational law (or legal computing) is a branch of law concerned with the study of formal representations and automated reasoning with laws (governmental regulations, business rules, and contracts) in electronically mediated domains. Like other disciplines in computational science, computational law is concerned with quantitative modeling and analysis techniques, e.g., by using computers to analyze and model legal issues. Many of the techniques used in computational law are taken or derived from techniques in the domains of natural language processing and big data analysis. Analysis techniques also include legal visualization techniques.

The present study is a part of main research to create encoding of law for electronic interchange of legal data. The law and procedures enshrined in the right to information rules 2012 has been transformed in encoded form as standard tags used as part of framework of codification. The encoding were done on framework which is XML based inspired by the AKOMO NTOSO schema [4]. However, the study kept its own simple schema, but pattern is AKOMO NTOSO.

Akoma Ntoso which means "linked hearts" in Akan language of West Africa provides a set of simple, technology-neutral electronic representations of legal/judiciary documents for ICT services in a worldwide context and provides an enabling framework for the effective exchange of "machine readable" legislations and other documents. [4]

eXtensible Markup Language(**XML**) describes data in a structured format using mark-up tags to distinguish data

elements. It is a standard specification of the World Wide Web Consortium (W3C) accepted way back in 1998. It provides a standard way to store almost any kind of data in a form that applications running on any platform can utilize. Unlike the pre-defined tags in Hyper Text Markup Language (HTML), mark-up tags in XML are not pre-defined. We can define our own tags just by adhering to specified naming conventions. The XML has many advantages such as **Readability** (documents are written in readable plain text), **Availability** (XML document can be created in plain text editor like Notepad), **Interoperability** (XML is W3C standard not reliant on any particular OS or other technology), Extensibility (new elements can be created and backward compatibility will be ensured by schema constraints), Ubiquity (XML parser exists on every platform and all produce same result), Adaptability (data can be easily added to an existing XML document by creating a new instances of existing elements). The XML is in immense use in applications like Really Simple Syndication (RSS), Mathematical Markup Language (MathML), Open Document Format (ODF), Scalable Vector Graphics (SVG), Rich Internet Applications(RIA) like AJAX techniques.

Akoma Ntoso stands for *A*rchitecture for *K*nowledge-*O*riented *M*anagement of *A*frican *N*ormative *T*exts using *O*pen *S*tandards and *O*ntologies. It is an international technical standard for representing executive, legislative and judiciary documents in a structured manner. It is a legal XML vocabulary and suggests also a naming convention for providing unique identifier to legal sources based on FRBR model.[5]

Right to Information Rules 2012 [6]- In exercise of the powers conferred by Section 27 of the Right to Information Act, 2005 (22 of 2005) and in supersession of the Central Information Commission (Appeal Procedure) Rules, 2005 and the Right to Information (Regulation of Fee and Cost) Rules, 2005, Right to Information Rules, 2012 has been created to deifne procedure to obtain information as per RTI Act 2005. RTI Act 2005 is a act to provide for setting out the practical regime of right to information for citizens to secure access to information under the control of public authorities, in order to promote transparency and accountability in the working of every public authority, the constitution of a Central Information Commission and State Information Commissions and for matters connected therewith or incidental thereto. The RTI act enables citizens of India, who desires to obtain any information under this Act and the Rules 2012 has been providing way to obtain information.

2.   OBJECTIVE
The objective of study was to add on the research towards the encoding of Indian law by creating of standard representation to be processed and read by the other systems for implementing electronic interchange of data (IED) through independent open format encoding of RTI Rules 2012. The encoded rules will be used for simulation of further research to achieve broader objective.

3.   RESEARCH METHODOLOGY
The study is related to create a standard representation of RTI Rules 2012 by encoding in open format like XML for interoperability and electronic exchange of legal data in systems and methods.In the absence of adequate background work a prototype simulation research design [7] was chosen to complete the encoding of RTI application and first appeal.

The encoding was done by creating XML documents and XSD schemas for RTI rules. The pattern was customized based on AKOMO NTOSO framework [4]. The primary data was the RTI Rules 2012 as promulgated by the parliament of India attached as ANNEXUE-I.

4.   DATA ANALYSIS, FINDINGS AND DISCUSSIONS
4.1   DATA
The Right to Information Rules 2012 consists of 15 Sections and 1 (one) Appendix. There are no separate chapters. Sections 1, 2, 4, 6, 8, 11, 12, 14 are divided into sub sub sections at one level. Section 10 of the RTI rules 2012 is further subdivided in sub sections in two level that means sub sections are divided into further sub sections. The Appendix was extension of Section 8 providing specification of format of appeal to Central Information Commission which termed as 2nd appeal. Format of appeal has 11 points but encoding of appeal has been kept out of this study. The word Section is used for rules and subsection for sub rules. The term section is used in all encoding to keep simplicity to refer the section, rule, article etc. unless otherwise defined. The structure of RTI rule 2012 created by classifying components of the rule. The detailed classification of structure component of RTI rule 2012 with xml encoding as schmea shown in Table 2 is shown in ANNEXURE-I of this paper. Summary of the RTI rules 2012 and XSD schema is depicted in below Table-1 & Table 2.

Table 1: Showing raw rules of RTI Rules

Sl	Col -1	Col-2
1	Type of Law	Rule
2	Derived From	RTI Act 2005
3	Total Number of Chapter	Single
4	Total Number of Sections	15
5	Total Number of Subsections	24
6	Total Number of Provsio	2
7	Section Heading	YES
8	Numbering Pattern	Irregular

Table-2: Showing XML Schema for RTI rules 2012 encoding in xml

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema           attributeFormDefault="unqualified"
elementFormDefault="qualified"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
```

```
<xs:element name="rtirule">
<xs:complexType>
<xs:sequence>
<xs:element name="meta">
<xs:complexType>
<xs:sequence>
<xs:element name="Name" type="xs:string" />
<xs:element name="GazzeteDate" type="xs:string" />
<xs:element name="Place" type="xs:string" />
<xs:element name="Country" type="xs:string" />
<xs:element name="Language" type="xs:string" />
<xs:element name="PublishedBy" type="xs:string" />
<xs:element name="ActNo" type="xs:string" />
<xs:element name="LegislationType" type="xs:string" />
<xs:element name="ActType" type="xs:string" />
<xs:element name="Govt" type="xs:string" />
<xs:element name="MainAct">
<xs:complexType>
<xs:sequence>
<xs:element name="ActTitle" type="xs:string" />
<xs:element name="ActNo" type="xs:string" />
</xs:sequence>
</xs:complexType>
</xs:element>
<xs:element name="PromulgatedBy" type="xs:string" />
</xs:sequence>
</xs:complexType>
</xs:element>
<xs:element name="preamble" type="xs:string" />
<xs:element maxOccurs="unbounded" name="section">
<xs:complexType>
<xs:sequence>
<xs:element name="num" type="xs:decimal" />
<xs:element name="heading" type="xs:string" />
<xs:element name="content">
<xs:complexType mixed="true">
<xs:sequence minOccurs="0">
<xs:element                maxOccurs="unbounded"
name="subsection">
<xs:complexType>
<xs:sequence>
<xs:element name="num" type="xs:string" />
<xs:element name="content">
<xs:complexType mixed="true">
<xs:sequence minOccurs="0">
<xs:element minOccurs="0" maxOccurs="unbounded"
name="clause">
<xs:complexType>
<xs:sequence>
<xs:element name="num" type="xs:string" />
<xs:element name="content" type="xs:string" />
</xs:sequence>
<xs:attribute name="id" type="xs:string" use="required"
/>
</xs:complexType>
</xs:element>
<xs:element minOccurs="0" name="provision">
<xs:complexType>
<xs:sequence>
```

```
<xs:element name="content" type="xs:string" />
</xs:sequence>
<xs:attribute name="id" type="xs:string"  use="required"
/>
</xs:complexType>
</xs:element>
</xs:sequence>
</xs:complexType>
</xs:element>
</xs:sequence>
<xs:attribute name="id" type="xs:string"  use="required"
/>
</xs:complexType>
</xs:element>
</xs:sequence>
</xs:complexType>
</xs:element>
</xs:sequence>
<xs:attribute name="id" type="xs:string"  use="required"
/>
</xs:complexType>
</xs:element>
</xs:sequence>
</xs:complexType>
</xs:element>
</xs:schema>
```

4.2   RTI RULES 2102 SCHEMA OBJECT MODEL

Following objects had been identified before creation of XML Schema

(A) Meta

Meta describes data about the law that is to be encoded. It may be termed as about the act. Following attributes has been identified for the RTI rules : (i) Name (ii) GazzeteDate (iii) Place (iv) Country (v) Language (vi) PublishedBy (vii) ActNo (viii) LegislationType (ix) ActType>Rules (x) Govt (xi) MainAct (xii) ActTitle (xiii) ActNo (xiv) MainAct (xv) PromulgatedBy

(B) Preamble

Preamble of the RTI rule describes its origin and purpose.

(C) Section

The Act contains 15 Sections under following headings (1) Short title and commencement, (2) Definitions, (3) Application Fee, (4) Fees for providing information, (5)   Exemption from Payment of Fee, (6) Mode of Payment of fee, (7) Appointment of Secretary to the Commission, (8) Appeal to the Commission, (9) Return of Appeal, (10) Process of appeal, (11) Procedure for deciding appeals, (12) Presence of the appellant before the Commission, (13) Presentation by the Public Authority, (14) Service of notice by Commission, (15) Order of the Commission

(D) Sub-section

(1) Short title and commencement has two subsections (2)

Definitions contains six subsections, (4) Fees for providing information has seven subsections, (6) Mode of Payment of fee has three subsections (8) Appeal to the Commission also consists of six sub sections (10) Process of appeal is divided into three sections, Sections (11) Procedure for deciding appeals contains six subsections, Sections (12) Presence of the appellant before the Commission has three sub sections and Section (14) Service of notice by Commission cosists of four sub sections

(E) Provision
Provision is an additional paragraph qualifying the law laid down in previous paragraph. In the said act only one provision was found at Section (10) Sub section (1) qualifying the law laid down in Sub Section (1) of Section (10) of Process of appeal.

(F) Clause
Clause has been taken as further branching of law in a sub section. It can be further classified in sub clauses but in the act under study has no sub clause of a sub section further.

Section (10) Process of appeal Sub section (3) has two clauses (a) and (b) laying some additional law.

The specific object, attributes and schema in detail will be kept at http://www.ecourtindia.in (The portal privately registered on which planned to keep all research material and encoding) and namespace for the XML encoding.

4.3   ANALYSIS   OF   ENCODING VALIDATION AND PRESENTATION OF LAW FROM XML

The encoded RTI rules will act also input to ongoing study to encoding of various component of justice delivery system of RTI Act 2005. The XML encoded RTI rules 2012 was tested for correctness using Xpath/ Xquerry statement through XSLT encoding. The validation Schema of xml encoding is depicted in Table 2 above. The XSLT querry to see result is shown in Table 3 below and result of the querry in Table 4 below:

Table 3: Showing XSLT code for RTI rules 2012 encoded file output

```
<?xml version="1.0" encoding="utf-8"?>
<xsl:stylesheet                          version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:template match="/rtirule">
<html>
<body>
<b>Total    Section    Count    :</b>    <xsl:value-of
select="count(/rtirule/section)" />
<br />
<table border="1">
<tr>
<th>Section No</th>
<th>Section Title </th>
<th> Sub Section</th>
```

```
</tr>
<xsl:for-each select="section">
<tr>
<td>
<xsl:value-of select="num" />
</td>
<td>
<xsl:value-of select="heading" />
</td>
<td>
<xsl:value-of select="count(content/subsection)" />
</td>
</tr>
</xsl:for-each>
</table>
<br />
</body>
</html>
</xsl:template>
</xsl:stylesheet>
```

Table 4: Showing XSLT execution Output

Section No	Section Title	Sub Section
1.	Short title and commencement -z	2
2.	Definitions -	6
3.	Application Fee -	0
4.	Fees for providing information -	7
5.	Exemption from Payment of Fee -	0
6.	Mode of Payment of fee -	3
7.	Appointment of Secretary to the Commission -	0
8.	Appeal to the Commission -	6
9.	Return of Appeal -	0
10.	Process of appeal -	3
11.	Procedure for deciding appeals. -	6
12.	Presence of the appellant before the Commission -	3
13.	Presentation by the Public Authority -	0
14.	Service of notice by Commission. -	4
15.	Order of the Commission -	0

Above output shows that RTI rules that has been encoded is producing the correct output regarding its content. XML encoded file has been tested for its contents by using different set of queries also and compared with original text. The CSS based formatting for presentation also compared with original act and found correct.

5.   SUMMARY, CONCLUSIONS AND RECOMMENDATIONS
5.1   SUMMARY
The objective of the study to provide add on to the research towards achieving standard representation of law using XML technologies. The successful encoding of RTI rules 2012 comparatively simplest form of law will provide a base to augment further study on encoding Right

to Information Act 2005. The xml schema created on the pattern of AKOMA NTOSO has worked well and a small but effective framework for implementing electronic exchange of legal data has been tested.

The Schema model of RTI rules 2012 which provides rules for implementation of RTI Act will support the real work of encoding RTI Act 2005.

In addition after standard representation of Right to Information act, the decision of various courts can be also encode and will result in natural language processing of law and its implementation for automated processes by achieving interchange data and understanding.

Use of XML extended the scope of objective and the accuracy above of 95% and at many places 100% provides way for complete encoding for RTI Act.

5.2   CONCLUSIONS

In context of legal system and processes in India the encoding of law under a standard representation will result in electronic exchange of legal data. The delivery of justice will get faster. The AKOMA NTOSA is at vey advanced stage needs to be worked in area of implementation and interpretation of law encoded. The output of queries on xml data of encoded RTI rules 2012 will provide the layout of framework of actual implementing the RTI Act 2005 encoding. The rule so encoded and application and appeal under the rule qualified by RTI Act is taken in another study, where the currently encoded RTI rules 2012 will act as input and new encoding of components of RTI implantation.

5.3   SCOPE OF FUTURE RESEARCH

This research is directed on the of implementation of Right to Information rules 2012. The result of the study makes pathe of further work in creating of standard representation for entire Right to Information Act 2005 and related rules, judgements, and other parts of the Act meant to be implemented for feasible use and further for representation of entire legal document.

6.   ACKNOWLEDGEMENT

In doing this research, many persons were part of the research in collection data, its analysis, etc. I here specially acknowledge the services of Computer Facility, RTI Cell, Academic Section.

7.   REFERENCES

[1] S. C. O. INDIA, "Supreme Court Judgments," [Online]. Available: http://sci.gov.in/judgments.

[2] e. o. e. India, "Ecourt Services of India," [Online]. Available: http://services.ecourts.gov.in/ecourtindia_v5/main.php.

[3] Indiankanoon.org, "Indian Kanoon - Search," [Online]. Available: https://indiankanoon.org/.

[4] A. D. I. M. P. F. V. -. D. C. U. o. B. Luca Cervone, "Document Architecture," [Online]. Available: http://www.akomantoso.org/akoma-ntoso-in-detail/document-architecture/.

[5] wikipedia.org, "Akoma Ntoso - Wikipedia," [Online]. Available: https://en.wikipedia.org/wiki/Akoma_Ntoso.

[6] GOI, The right to information act, 2005 No. 22 of 2005, Commercial Law Publishers (India), 2005.

[7] C. R. Kothari, Research Methodology - Methods & Techniques, New Agea International Publishers.

[8] S. NAIB, The Right to Information Act 2005, A handbook, New Delhi: Oxford University Press, 2011.

[9] H. D. P.J. Dietal, Internet & World Wide Web, Pearson Prentice Hall, 2009.

[10] D. E. Services, Html 5 Black Book, Dreamtech, 2017.

❑❑❑

BETTER TECHNOLOGY BETTER WORLD : A SURVEY

[1]Ms. Preeti Johri and [2]Dr. Anuranjan Mishra
[1]Research Scholar, Department of Computer Science & Engineering,
Noida International University, G.Noida, Uttar Pradesh, INDIA.
[2]Prof. & Director, Accurate Institute of Management and Technology, G.Noida, Uttar Pradesh, INDIA.
*Corresponding Author Email ID: preeti.johri002@gmail.com

ABSTRACT

This paper highlights main issues emerged when the implementation of ICT based learning will be done in the Government schools. As Government is funding a big amount to make the country educationally sound still there are some teachers who do not want to upgrade themselves and they do not even ready to change their old teaching strategies. So, in this paper we will discuss the techniques which the science teachers of Government schools following these days. As Government organizes so many seminars and workshops for the teachers every year still they are not upgraded. They don't include ICT in their daily lectures. They think they don't have any need to include computers and other technologies in their teaching. At today's time we see everything is getting online if a person doesn't have the knowledge of computer others make fun of him/her. This is because IT is the necessity of our society.

Keywords: Seminars, Workshops, Classification, Clustering, WEKA, IWB, Animation, Feedback, Interactive, Spotlight, Digital, TMT, SCT.

1.   INTRODUCTION

School education is a concurrent subject under constitution of India and falls within the domain of Center and States/U.Ts. In the rural areas also, the Government play a significant role in providing scholarships, free-education to the girls, and so on. School Education is a vast domain with 1,29,800 secondary schools and over 11 Lakh elementary schools spread access the country and employing more than 55 Lakh teachers. Most of the states have already taken initiative for introducing the ICT at school level. A National Policy on ICT in School Education has been recently finalized by the MHRD to provide broad framework to guide the states for the use of ICT tools holistically in schools. Some NGOs (like Aziz Premji Foundation) are also there who emphasize these days on the usage of ICT in the education sector. Some IT companies like Aptech, Averonn are also providing IT training in the Government schools to make the teachers there computer trained. NCTM focuses on technologies to teach mathematics in school. These technologies are used these days for interactive and effective learning. Technology-oriented learning is now available in many well developed countries today to enhance effective learning in students. Therefore, the new teaching strategies must have to acquire all effective teaching and learning materials that accommodate different student's requirement and interest. The smart school project started in Malaysian Ministry of Education followed by Iran Ministry of Education [2]. As per Government officials, basically high performance achievement of students boosts our economic growth and get ready for the competitions

(European Commission, 2010, 2011)[3]. Language practices and Mathematics are the basis of the student's growth, as some reports have already shown that most of the students not having the sufficient essential knowledge about mathematics and language practices. We can inculcate these skills in our students with fueling of Information Tools. IT techniques are appreciated to enhance student knowledge, they can be used it at home as well as at school. It is of no surprise if schools are introducing these digital devices in their education system like computers, laptops, interactive whiteboards, iPads etc.[7] to increase student knowledge and interest. According to the survey, some stated that there will no benefits of ICT on education, some agreed with positive benefits of it and some said there are negative effects of IT techniques on the students' career [3]. IWBs are used to integrate the functions of a regular class board + additional functionalities means to give the input to the computer perform the calculations and manipulations on the data; IWB enable learning and teaching more interactive and constructivist. The IWB is a technology combines a computer connected to both a projector (to highlight the things) and a touch-screen board (to take input from user, allows changes). Its main features are: Drag and drop, Hide and reveal, Highlighting, Animation, Storage and recall, and Feedback.

Smart class was a digital initiative of EDUCOMP (the smart class creator), which is rapidly changing the way of teaching and learning among all the participants. The Smart Class software has brought a drastic change and

transformation in the methods of learning. The system fulfills all the student's requirement and provides innovative learning solutions in an easy way using digital instructional tools [5].

2. METHODOLOGY

For this study, the researcher prepared a questionnaire for the students of XII standard of Government schools where they have a lot of pressure and have to emphasize on their results. One questionnaire is for the teachers:

- *Sample:* Here we collected the data from XII standard students and the teachers of school. We distributed the questionnaire among the XII students and the teachers for just ticking the answers of their choice. Questionnaire included some simple questions so that it would be less time consuming.
- *Tools:* We analyze the data on the basis of

questionnaire. First we prepared the questionnaire then we took the permission from the authorities to conduct the survey[6]. The questionnaire was categorized into two- (i)For teachers and (ii)for students.

First we made the data list for the relevant answers then we applied preprocessing tools on the data using WEKA tools. WEKA tools are the data mining software which extracts the data from a list and presents it in a manner.

3. DIFFERENT TEACHING METHODS

There are different teaching methods we follow in our education system: Smart Classroom teaching (SCT) and traditional mode of teaching (TMT). [8]

Table 1: Showing Distribution of Sample on the basis of different classroom teachings

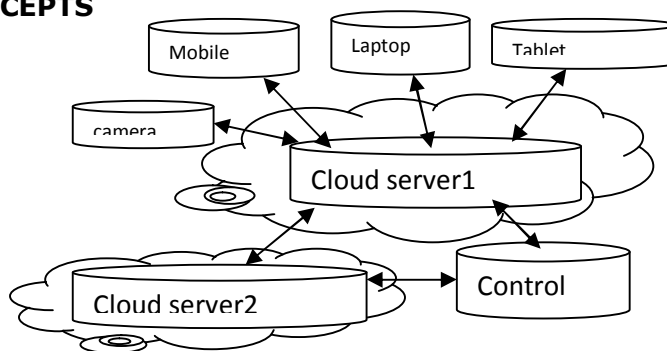| S. No.	CLASS	GROUPS		Total No. of STUDENTS
		Experimental (Group LIKED learning with SCT)	Control (Group LIKED learning with TMT)	
		No. of Students	No. of Students	
1.	X	120	40	160
2.	XI	40	70	110
3.	XII	45	85	130
Total		205	195	400

4. STORAGE CONCEPTS

Fig. 1: Storage Concepts

Above fig. showing the communication between various types of devices to read the data from the server and to store the data on the server. There are some nodes also working with server to make the data secure.

5. ALGORITHMS

Step 1: Start
Step 2: Initialize A[] with some message
Step 3: For A[n] where n= 1 to n-1
Step 4: If A[n]<n-1
        Then Replace A[n] with some letter
Step 5: Read the string

The string will be encrypted with some letters like, we

want to encrypt the text replacing the I with W letter. The text will be decoded at the destination system.

Algorithms will work on databases for updating the data and providing the authentication to the persons who can access the database and at what extent. So that our data will be consistent throughout.

6. KEY DRIVERS

Some of the key drivers identified during the design of this study are provided below:

- Improve the teaching methodologies
- Assess the students online
- Conducting online exams for each and every subjects

like computer, science, s.st and mathematics also.
- Take the education system at high reach
- Give the flexibility to the students to learn practically.

7.   RESEARCH FINDINGS

Whenever we use digital devices to store the data then there always be a chance for that data to get hacked. So we must take care about the algorithms can be applied to make our data secure. Then it will only be possible to make our world fully techno oriented.

8.   OBJECTIVES

The objectives of this study are:
- Improving quality of learning
- Make efficient school administration and governance
- Improved and better service delivery of school education department to the students, parents, community, teachers
- To find out a summary from the data sets related to the questions about IT enabled learning and teaching
- Necessary points to know about Technology oriented assessment

Table 2: Learning Support Services: ICT Enabled Training, Teaching, and Assessment Services

S. No.	SERVICE
1	Content Platform Services to create the portal for hosting, accessing Student and Teacher Resources.
2	ICT enabled Teacher Training and classroom teaching Services.
3	Standard Assessment Services for teachers and students (SSA & States initiatives)
4	Self-Learning Tools for teachers and students.

9.   IMPLEMENTATION COMPONENTS

The implementation components factored for the project costing are listed below:
- Implementation of ICT Solutions and Digital

- Learning Techniques to appreciate the services identified for betterment
- Hosting Services, compute and storage
- infrastructure, and Bandwidth at the Data Center
- Change management and training procedures for these changes.

10.   CONCLUSION

As we know that our society is technology based these days, we have to put more efforts in this direction. Then only, we will get the better and enhanced world.

11.   REFERENCES

[1] Arora, P.,The ICT laboratory: Analysis of computers in public high schools in rural India, AACE Journal, 15(1), 57-72. (2007)
[2] Hamzah, M.I, Embi, M.A., Ismail, A. ICT and Diversity in Learners' Attitude on Smart School Initiative, International Conference on Learner Diversity 2010.
[3] Sofie J. Cabus12, Carla Haelermans3, Sonja Franken4 ,SMART in Mathematics? – Exploring the Effects of In Class Level Differentiation using SMARTboard on Math Proficiency, British Journal of Educational Technology October 2015
[4] Tal Berger Tikochinski and Rachel Zorman Henrietta Szold Institute, Israel, Using the Interactive White Board in Teaching and Learning – An Evaluation of the SMART CLASSROOM Pilot Project, IJELLO special series of Chais Conference 2011 best papers
[5] Inderpreet Kaur Chachra ,Effect Of Smart Classroom Assisted Teaching On Academic Achievement Of Students Of Different Intelligence Level In Social Science , Abhinav National Monthly Refereed Journal of Research in Arts & Education , Volume 4, Issue 6 (June, 2015)
[6] Preeti Johri and Dr.Anuranjan Misra, "Government schools focussing on ICT based education- A Survey, ICDMAI 2018, PUNE.
[7] Preeti Johri and Dr.Anuranjan Misra, "Government schools focussing on ICT based education- A Survey, BVI, INDIACOM 2018.
[8] Preeti Johri and Dr.Anuranjan Misra, "Technology in K-12 Education: A Boom or Bane, MAT Journal, Journal of Web Development and Web Designing Volume 2 Issue 2

❑❑❑

PROCESSING CHALLENGES OF MAPREDUCE INTO CLOUD ENVIRONMENT

[1]Suyash Mishra*, [2]Dr. Anuranjan Misra and [3]Dr. Suryakant Yadav
[1]Computer Sciences Department, Noida International University, G.Noida, Uttar Pradesh, INDIA.
[2]Prof. & Director, Accurate Institute of Management & Technology, G.Noida, Uttar Pradesh, INDIA.
[3]Computer Sciences Department, Noida International University, G.Noida, Uttar Pradesh, INDIA.
*Corresponding Author Email ID: ssuyashmishra@gmail.com

ABSTRACT

MapReduce is a programming model, proposed by Google initially to process large datasets. It consist of two components Map and Reduce motivated from functional languages e.g. Lisp and Haskell. To achieve desirable result programs task is to define and implement Map and Reduce functions. Map and Reduce should describe the logic encapsulated in both components to achieve desirable result. There are various big data based operations e.g. culling, highlighting, indexing, searching, faceting. It impossible for a single node machine to manage and process such a huge data. Map Reduce regarded as an important function to tackle continuous increasing demands of computing resources required for processing big data sets. MapReduce highly scalability nature permits parallel and distributed processing on multiple computing nodes. This paper talks about MapReduce usages in cloud computing and motivate its use for big data processing into cloud heterogeneous environment.

Keywords : Big Data, Iot, Cloud Computing, Grid, MapReduce.

1.   INTRODUCTION

Traditional data storage and processing capabilities were limited and was dependent on underline hardware configurations, storage and processing requirements, which deemed to be very different from today. Due to increase in data, volume and unstructured data traditional methods and databases are facing huge threat to ingest and process Big Data processing and storing demands. Now a days Industry is focusing and making huge investment to conclude how to make better use of Big Data and identify beneficial business insights to lead a better business decisions .Which help business or industry to increase profit. MapReduce is a highly scalable programming model able to process huge volume of data in parallel execution fashion on a huge number of commodity computing nodes. Google [3] has developed MapReduce paradigm was later used and implemented in many open source projects, but the Apache Hadoop harnessed processing most efficiently.

Now a days Industry is focusing and curious to know about efficiently utilize Big Data and analyze to identify beneficial business insights for making them quipped with viable and better business decisions .Which in turn add value to their business ,increase profit ,retain customers and make customers happy with services they are offered.

Advancement in the technology has also brought flood of unstructured data. Analysis of this data is important to extract value from the abundance of data available. The rate at which the data is growing is very high and unpredictable. Businesses can use this data for multiple purpose. This data is further utilized for conducting customer trend analysis, customer's sentiment analysis or to know customer feedback on service or product Company is offering. Which can be further converted into structured data-by-data analyst as per their convenience.

Big data is comprised of huge volume of structured, semi-structured and unstructured data that used for data analytics using various methods.

2.   HADOOP

Hadoop is an open source-processing engine designed to compute extremely large unstructured or semi structured datasets efficiently. Hadoop follows distributed processing mechanism, which offers resilience and scalability while processing big data. Primarily Hadoop has two major constituents HDFS and MapReduce. HDFS (Hadoop distributed file system) manages data inspired from UNIX file system responsible for storage management in structured relational form or unstructured form and in any form in between similarly. HDFS is a highly distributed file system ensures highly computation of big data along with scalable facility. The MapReduce programing model manages applications on multiple distributed servers or

user related processing task in an efficient manner. This provides an environment to execute a highly efficient data processing by harnessing distributed processing feature, which is based on divide and aggregate paradigm. HDFS gives liberty to user or application programmer to capitalize unlimited storage need. Figure1 below depicts both prime components of Hadoop framework.
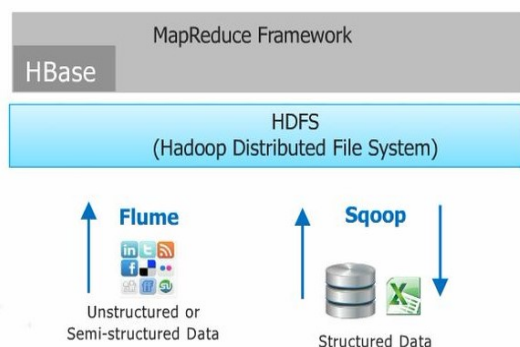

Fig. 1: Hadoop compnents

3. PROCESSING OF BIGDATA USING MAPREDUCE

In the MapReduce paradigm, MapReduce [2] has become the prominent batch-processing model. MapReduce is a very flexible programming method able to parallel process large volume efficiently on large number of computing nodes.

MapReduce methodology consisting of map and reduce functions, which allows application developers to customize them as per their requirement. Along with providing faster processing Map Reduce manages node failure and abstracting design complexity from the programmer.

MapReduce has been developed for effective parallel processing of big data by breaking the load into independent sub units. Prominent benefit of MapReduce is that it conceals system level implementations and complexities from developer by allowing developer just to focus on its objective. MapReduce is a two-phase approach: Map and Reduce. Every Map phase takes input files present into various nodes of distributed file system. If Map function does, co- located with data partition system will attempt to move data portion to node where Map function exists with data portion to reduce data movement. Hence MapReduce backs concept of as "Moving data closer to compute" to optimize execution time. Post Map function finishes its processing, reduce function is then further run on all values having same interim key value and output key/value pairs as the result. The MapReduce framework based on master/slave architecture having a one node acting as master has Job Tracker and several slave nodes, which runs assigned task, and Tasktrackers, one per node in the cluster, maintain their status. The Job Tracker acts as communicator between users who submits the job and the underline pool

of hardware. Job Tracker accepts users assigned map/reduce jobs to the and further jobs are executed on sequence of submission first come/first-served basis. The Job Tracker does resource management required for map, reduce steps, allocated, and monitor slave nodes executing allocated task via tasktrackers. The Tasktrackers accepts and run tasks allocated by JobTracker, manages data movement among the map, and reduce step.

Whole processing can be elaborated in detail as per below processing sequence it has been explained in figure2.

- *Map* – Input data is first accepted by master node, which divides data file into smaller datasets, and move them to slave nodes. A slave node may reiterate the process leading to a multilevel tree structure. Map accept similar type of data and gives a list of output.
- *Map logic execution* – Map logic will run once for every key value and generates output ordered by key values.
- *Reduce phase execution* – the MapReduce system identify a node to run Reduce function, assigns the key value to each processor, and provides that node with all the Map-generated data allocated by same key value.
- *Execute Reduce code* – Reduce is executed only once for each key value produced by the Map stage.
- *Produce the result* – Final output is generated by consolidating all the Reduce output, and sorts them by key.

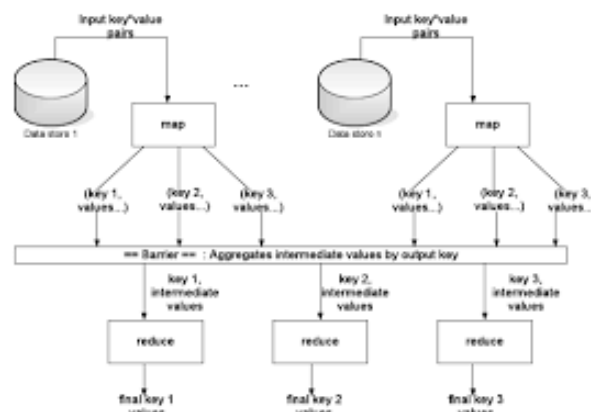Below diagram Figure2, elaborates word count example of MapReduce processing.

Fig. 2: Steps of Mapredcue Execution

In a homogeneous environment, all nodes have equal computing speed and disk capacity. In case of node, failure happens MapReduce attempts to reruns failed tasks on a different node, which is idle/less-occupied at that moment. In some instances a node is executing a task very poorly or slowly known as straggler task, in that case Map Reduce kick-off a speculative copy of slow running task ("backup task") on another node, which can finish task quickly job will be as slow as the misbehaving task ,if no speculative task is there. Straggling tasks can also be triggered due to impaired hardware and slowness This can

improve job processing by 44%.In Heterogeneous environment lack of proper mechanism to address speculative task may increase processing time due to difference in nodes storage and processing capability.

4.   MAPREDUCE: PROCESSING OF DESKTOP GRID COMPUTING

Grid computing assumed to be the combination of computer resources spanned across multiple places with a single objective. The grid is a distributed system consists of non-interactive nodes involving a number of input data files. Grid computing is different from traditional high executing computing nodes in such a way that grid computers each node executes a different task or application. Grid nodes are diverse hence; form more heterogenenity.A single grid can be assigned to an application having similar processing requirement.

Foster and Kesselman [9] has defined that desktop Grids are systems that combine the computation capability of idle nodes, taking benefits of the fact that these machines are available during most part of the day (85% during day time, in addition, as much as 95% during the night [9]. This environment also comes with verity of concepts.

Heterogeneity .There is diverse set of hardware and software configurations that affects the processing capability of individual nodes and the performance of the grid as a whole. Therefore, less occupied nodes are assigned tasks to process grid tasks. As additional aspects that can help to characterize this environment, we could cite:

There are below problems running MapReduce deployments on desktop grids in comparison homogeneous environments.
- More tasks/job failures probability, when users interrupt the execution of nodes.
- Distributed file system nodes will also become unavailable may cause missing data besides task failure.
- Excessive launch of speculative tasks along with data distribution due to heterogeneity of nodes.

5.   MAPREDUCE APPLICATION IN CLOUD FOR BIG DATA PROCESSING

Cloud computing is an advance programming suited for meeting parallel and distributed computing. It consists of interconnected and virtualized computers network. With its advancement, new MapReduce implementations developed to deploy and run MapReduce in cloud environment.
- Cloud services are different from normal services in below manner.
- Based on "PAY AS YOU GO", means can be sold for hour's days or monthly usages.
- Elastic in nature, wherein users can dynamically change demand to get more or less services as much as they need.

The key technical challenges identified on cloud systems are as below:

Virtualization: Virtualization is a kind of making a virtual image of a server or services e.g. operating system, storage media or networking resources having an intent to be used across multiple machines simultaneously. Prime objective of virtualization is to take care of workload management by redefine traditional computing by optimizing the processing, efficient and economical Virtualization technology is hardware optimizing cost effective technology that is rapidly advancing the fundamental way of computing. Hence maintaining efficient virtualization is key for Cloud based systems.

Multitenancy: A phenomena in cloud systems, where the code location or data location is not known and copy of same resource is allocated to other users known as tenant and these phenomena known as multitenancy. This makes software managing easy and cost effective by sharing resources, cost .This makes high availability of resources and services, which are hosted on shared resources are must be always available for other users. Data Management is an important aspect for storage clouds, where data distribution across various resources is done it self.

Security: Data security, Compliance and privacy is an extremely important feature in cloud system to protect sensitive data and code. Main challenge for MapReduce in cloud is load balancing .In a traditional Hadoop there are two nodes name (Master node)node and data nodes(slaves) .However in cloud environment this name node is a cloud server and slaves nodes may also be distributed server. Data files are already loaded into cluster nodes. When the MapReduce starts, execution NameNode picks up JobTracker to allocate tasks, TaskTracker monitor which DataNodes and progress will finally process. There are many Map programs, which can execute on each Data Node and the interim results will be send to next process, which combines and produces result.

Load balancing: It plays vital role in distributing the load uniformly across the available /underutilized nodes. When a node became over utilized than what it can process .Though load balancing load is shared to available nodes is not so important in a MapReduce processing, it becomes essential to process large files when hardware resources comes at a cost and usages is critical. Hence, load balancing optimizes hardware optimization in resource-critical scenarios with a significant uplift in performance.AWS EMR allows user to run and submit their jobs via web services rather than taking traditional steps to run over Hadoop makes analyst life easy. This offers immense opportunity to dig more onto unstructured big data generated over cloud EMR merits e.g. reliability, convenient to use, cost effective, scalable, and secure in

nature makes this best choice in cloud environment.

There are below areas where Amazon EMR can prove instrumental to benefit analysts and industry.

Clickstream Analysis: Click stream analytics empowers user to collect, process, takeout summarized outlook of website visitor .This helps organizations to know user segment, user click behavior and based on analysis appropriate, and effective advertisements can be sent to user. This helps organization to know their customers better and offer product of their need make customers more satisfied and retain them with business. Amazon EMR is suitable and used for click-stream data analytics.

Real-time Analytics: Real time data flowing into system require quick storage and process real-time data coming from various sources .This help business to know real time behavior of users and offers services accordingly. There are many tools and vendors e.g. Amazon Kinesis, Apache Kafka, and many more data streams for managing and respond the real time data .For instance Spark Streaming is famous for real time data processing in quick manner and then write processed result back to Amazon S3 or HDFS.

Log Analysis: Amazon EMR helps organization process logs generated by web applications, sensors data and mobile applications. Amazon EMR converts petabytes of un-structured data into semi-structured or structured data so that analytical operations to be performed to derive useful insights about customer's behavior or application functionality.

The AzureMR also provides cloud based MapReduce facility using the HDInsight .A MapReduce job can be submitted using web services, monitor its progress and store output to HDFS for further computation and analysis. The main contribution of AWS and Azure computing is a new architecture, without a server node as below.

6.   CONCLUSION AND FUTURE WORK

MapReduce has been regarded as prominent programming paradigm to cope with Big data processing .Though MapReduce offers numerous advantages but there are few trade-offs faced in meeting, the rapidly growing computing demands of Big Data in heterogeneous environment. There are many scheduling methodologies proposed .Our aim is to identify and categorize related scheduling algorithms, their capability to address MapReduce challenge to work efficiently in

Heterogeneous environment. .This enables better planning of Big data projects. Future work on this will be develop a scheduling change in Hadoop MapReduce which will work efficiently using LATE scheduling approach as this is described as most suited approach among all proposed scheduling methodologies. In addition, investigation will be done to compute finish time estimation in more detail.

7.   REFERENCES

[1] Dean, J., Ghemawat, S.: MapReduce: Simplified data processing on large clusters. Year 2004

[2] Hadoop MapReduce, http://hadoop.apache.org/mapreduce/

[3] "Towards Scalable Systems for Big Data Analytics" by H Hu, Y Wen, T Chua, X Li, year 2014

[4] Chang, V. Towards a big data system disaster recovery in a Private cloud. Year 2015

[5] Zhang, L. et al "Moving big data to the cloud" IEEE 2013.

[6] Wu, X. et al. "Data mining with big data" IEEE 2014.

[7] Subashini, S. & Kavitha, V., "A survey on security issues in service delivery models of cloud computing" published in. Journal of Network and Computer Applications 2011.

[8] Hashem, I.A.T. et al." The rise of "big data" on cloud computing" 2014.

[9] I.Foster and C. Kesselman. "The grid: blueprint for a new computing infrastructure." 2004.

[10] https://aws.amazon.com/emr/

[11] https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-use-mapreduce

[12] D. Kondo, G. Fedak, F. Cappello, A. A. Chien, and H. Casanova. "Characterizing resource availability in enterprise desktop grids".

[13] D. Kondo, B. Javadi, P. Malecot, F. Cappello, and D. P. Anderson. "Cost-benefit analysis of cloud computing versus desktopgrids." IEEE 2009.

[14] Case Study Nokia: Using big data to Bridge the Virtual & Physical Worlds. Geller , cloudera, 2012.

[15] J. Ekanayake, H. Li, B. Zhang, T. Gunarathne, S. H. Bae, J. Qiu, and G. Fox. " Twister: a runtime for iterative mapreduce". ACM 2010.

[16] M. C. Schatz. Cloudburst: highly sensitive read mapping with mapreduce. Bioinformatics, June 2009.

[17] T. White. Hadoop: The Definitive Guide. O'Reilly Media, original edition, June 2009.

[18] J.Xie, S.Yin, X.Ruan, Z.Ding, Y.Tian, J.Majors, A.Manzanares, and X. Qin. "Improving mapreduce performance through data placement in heterogeneous hadoop clusters" IEEE 2010.

[19] Mahesh, A. et al. Distributed File System For Load Rebalancing In cloud Computing. , 2, pp.15–20. ., 2014

[20] M.Zaharia, A.Konwinski, A.Joseph, R.Katz and I.Stoica. Improving mapreduce performance in heterogeneous environments.

❑❑❑

A COMPARISON BETWEEN CONTROLLER DESIGN USING CHEBYSHEV-POLE CLUSTERING APPROXIMANTS AND PADÉ APPROXIMATION TECHNIQUE

[1]Vartika Rao* and [2]Vijaya Yaduvanshi
[1,2]Assistant Professor, Department of Electrical, Electronics & Engineering,
[1]AIMT, G.Noida, Uttar Pradesh, INDIA.
[2]United Institute of Technology, Allahabad, Uttar Pradesh, INDIA.
*Corresponding Author Email ID: raovartika79@gmail.com

ABSTRACT

This paper presents comparison of two methods of designing controllers. First method for designing of a controller based on Chebyshev-Pole Clustering Approximation pertaining to unit step input. The designed controller reduces the intricacy of the higher order controller. The coefficients of numerator are determined by Chebyshev polynomial series and the denominator polynomial is obtained by pole clustering technique and the second one presents a method to commence a reduced order system for a given SISO (stable) linear continuous time system. This approach is about to retain the stability while converting the higher order system into its lower order approximant using Padé approximant and then a controller is put together in the reduced order system attaining stability and necessary parameters of the system.Numerical example is also presented to illustrate the behavior of the original system with its reduced order approximation and then attaching a controller and check the stability of the reduced order system with controller.

Keywords: Controller Designing, Model Truncation, Padé Approximation Technique, SISO Linear Continuous Time System, Chebyshev Polynomial, Pole Clustering Technique.

1. INTRODUCTION

In many cases, it is quite essential to illustrate a high order system by a lower order system. System reductions of continuous and discrete systems have been broadly examined. There are several techniques which are Aggregation method [14], Moment matching technique [15], Padé approximation [16], Routh approximation [17], L^{∞} optimization technique [18]. Padé approximation provides computational modesty and fitting of time moments.

But in many cases it provides instability in the reduced order model while the early system is stable. To get a truncated order system; Shamash [19] has already provided a technique of merging the Routh approximation and time moment matching. Just have a glimpse [20] of the technique in which denominator of the system is taken by keeping the dominant poles of the system and the numerator is achieved by comparing the time moments. These techniques provide certainty in obtaining a truncated order system and these are often called Partial Padé approximation in the frequency domain. A time domain version of these Padé approximation techniques have been described by Bandyopadhyay and Lamba [21]. A consolidation of frequency and time domain Padé approximation is also illustrated in [22].

Reduced order control modeling techniques, Anderson and Liu [23] are characterized in two types, direct and indirect techniques. In Direct techniques controller order is confined firstly and then find gain by extension, while indirect technique truncate the size of high order controller. Optimal projection theory, Gangsaas et al.[24]and Bernstein and Hyland are the direct techniques and the parameter optimization approach. A composite method based on Pole Clustering and Chebyshev Polynomial Technique of obtaining stable reduced order model have been proposed in 2012.

2. PADÉ APPROXIMATION TECHNIQUE

Let us assume the transfer function illustrate a stable single input single output (SISO) system which is given below:

$$G(s) = \frac{\Omega_1 s^{n-1} + \Omega_2 s^{n-2} + \text{---------} + \Omega_n}{s^n + \Gamma_1 s^{n-1} + \text{---------} + \Gamma_n} \quad (1)$$

$$= t_1 + t_2 S + \text{---------} + t_n S^{n-1} + \text{-----------} \quad (2)$$

(Expansion around s=0)
$$= M_1 S^{-1} + M_2 S^{-2} + \text{------} + M_n S^{-n} + \text{---------} \quad (3)$$
(Expansion around s=∞)

To drive its poise compressed order (rth. order) approximant, the transfer function will be:

$$G_{comp}(s) = (4) \frac{v_1 S^{r-1} + v_2 S^{r-2} + \text{--------} + v_r}{S^r + X_1 S^{r-1} + \text{--------} + X_r}$$

$$= \Delta_1 + \Delta_2 S + \text{--------} + \Delta_n S^{n-1} + \text{----------} \quad (5)$$

$$= N_1 S^{-1} + N_2 S^{-2} + \text{------} + N_n S^{-n} + \text{--------} \quad (6)$$

To acquire the Pade approximant, comparing the first 2r items of equation (2) to the 2r items of equation (5) respectively. Sometimes Padé approximant gives ambiguity in the response of the system. By introducing many stable reduction methods planted on the retention of r items this issue can be overthrown; have a look, for illustration [1]-[12].Many a times, it is not plenty enough to compare r items for a satisfactory total time response approximation[5];both time moments and Markov parameters must be taken. To maintain stability, a number of methods are in existence that helps in fully retaining r-items. Here according to the previous results Vimal Singh[13] by using Routh-Padé approximants it is viewed that, the denominator of the system must be taken, So as to reduce error between the (r+1)th and the consecutive time period. To retain stabilized system; Markov parameters are introduced where the numerator can be obtained in such a way, by fully restraining the first r-time moments /markov parameters of the system

3.   POLE CLUSTERING

The inverse distance measure criterion for calculating the cluster center from the poles of the higher order system is proposed by Sinha and Pal. Let r real poles in one cluster group and P_c be the cluster center based on IDM criterion

$$P_c = \left\{ \left(\sum_{i=1}^{r} \left(\frac{1}{hi} \right) \right) \div r \right\}^{-1}$$

Where P_c is cluster center of r real poles h_1, h_2, h_3, h_r of the original cluster.

Let r real pair of complex conjugate poles be $(h_{1R} + jh_{1I})$,

$(h_{2R} + jh_{2I}) \ldots (h_{rR} + jh_{rI})$

$$h_r = \left\{ \left(\sum_{i=1}^{r} \left(\frac{1}{hiR} \right) \right) \div r \right\}^{-1}$$

$$h_I = \left\{ \left(\sum_{i=1}^{r} \left(\frac{1}{hI} \right) \right) \div r \right\}^{-1}$$

While calculating the centers of cluster by inverse Distance measure Criterion, the cluster center for real and imaginary poles should have separate calculations the retention of the poles on imaginary axis should be done.

The denominator of the reduces model can be calculated by:

#1: Let all the calculated cluster centers are real,

denominator comes as:

$$\widehat{Dr}(s) = \prod_{i=1}^{r} [s - (P_C)_i] \prod_{i=1}^{r} [s - (P_C)_i]$$

= Denominator of the Reduced order Model.

#2: Let all the calculated cluster centers are complex conjugate, denominators comes as:

$$\widehat{Dr}(s) = \prod_{i=1}^{r/2} [s - (h_R + jh_I)][s - (h_R - jh_I)]$$

#3: Let one cluster center is real and the remaining one is complex conjugate, denominator comes as:

$$\widehat{Dr}(s) = (s - P_1) \prod_{i=2}^{(r+1)/2} [s - (h_R + jh_I)][s - (h_R - jh_I)]$$

4.   CHEBYSHEV POLYNOMIAL

The Chebyshev polynomial Tn(x) of the first kind is a polynomial in x of degree n, defined by the relation

Tn (x) = cos nθ when x = cos θ

If the range of the variable x is the interval $[-1, 1]$, then the range of the corresponding variable θ can be taken as $[0, \pi]$.These ranges are traversed in opposite directions, since $x = -1$ corresponds to $\theta = \pi$ and $x = 1$ corresponds to $\theta = 0$.

It is well known that $\cos n\theta$ is a polynomial of degree n in $\cos \theta$, and indeed we are familiar with the elementary formulae

$\cos 0\theta = 1$, $\cos 1\theta = \cos \theta$, $\cos 2\theta = 2 \cos 2 \theta - 1$,
$\cos 3\theta = 4 \cos 3 \theta - 3 \cos \theta$, $\cos 4\theta = 8 \cos 4 \theta - 8 \cos 2 \theta + 1$

Now, that the first few Chebyshev polynomials are:

$T_0(x) = 1$, $T_1(x) = x$, $T_2(x) = 2x^2 - 1$,
$T_3(x) = 4x^3 - 3x$, $T_4(x) = 8x^4 - 8x2 + 1$

5.   PRIMARY RESULTS

Considering the outcomes [13], we can have an idea that with the help of Padé approximation technique a higher order system can be reformed into lower order system retaining its stability. Here the closed loop transfer function of a system is given below:

$$G(s) = \frac{8s^2 + 6s + 2}{s^3 + 4s^2 + 5s + 2} \quad (7)$$

Equation (8) is attained by Routh-Padé approximation techniques; planted on the contemplation of time moments and it nearly preserves r+2 time moments. So equation (8) is the reduced order approximant which is given in equation (7).

$$G_{comp}(s) = \frac{8.0s + 8.129004}{s^2 + 4.307413s + 8.129044} \quad (8)$$

6.   CONTROLLER ARCHITECTURE

For a given control system; Fig.1 Gpri(s) and H(s) are already provided. Our main goal is to acquire the transfer function of the controller Cp(s) and with the help of Cp(s) desired response of the closed loop system is obtained. For modeling of controller Cp(s) an indirect technique is used here. To model and acquire the closed loop transfer function of the controller; assumptions for model specification of the compressed order model has been taken.
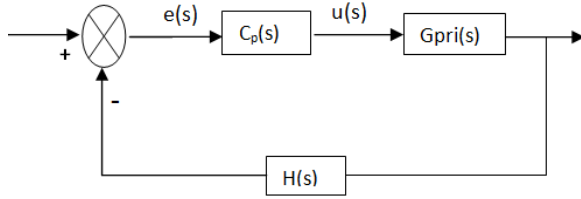

Fig.1: Control Structure

$$G_{fpp} = \frac{Cp(s)\,Gpri(s)}{1+Cp(s)Gpri(s)\,H(s)} \quad (9)$$

So,

$$G_{comp}|(s) = \frac{Cp(s)\,Gpri(s)}{1+Cp(s)Gpri(s)\,H(s)} \quad (10)$$

After modification for controller, the transfer function will be:

$$C_P(s) = \frac{G_{comp}(s)}{G_{comp}(s)[1-G_{req}(s)]} \quad (11)$$

With the help of Padé approximation technique $G_{pri}(s)$ can be easily estimated by a reduced order transfer function $G_{comp}(s)$ shown in Fig. 2

$$G_{fcp}| = \frac{Cpcomp(s)\,Gpri(s)}{1+Cpcomp(s)Gpri(s)\,H(s)} \quad (13)$$

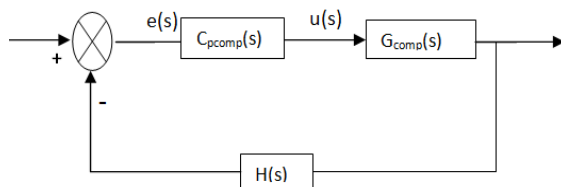To get a reduced order controller, $C_{pcomp}(s)$; the method has been explained.

Fig. 2: Closed Loop control with Cpcomp(s) and Gcomp(s)

7.   NUMERICAL EXAMPLE(PADE)

The transfer function of the primary system is [13]

$$G_{pri}(s) = \frac{8s^2+6s+2}{-3\,s^4\,+s^2\,+5\,s+2} \quad (14)$$

Let us consider a reference model. In this example, a standard second-order transfer function is taken with damping ratio ϵ =0.7 and natural frequency w_n=1.5 rad/sec. Therefore

$$G_{ref}(s) = \frac{\omega_n^2}{s^2 + 2\omega_n\epsilon s+\omega_n^2} \quad (15)$$

$$G_{ref}(s) = \frac{2.25}{s^2 + 2.1s+2.25} \quad (16)$$

A second order model given in the equation (17) is obtained by Padé approximation technique.

$$G_{comp}(s) = \frac{8.0s+8.129004}{s^2+4.307413s+8.129044} \quad (17)$$

Now let us calculate the transfer function of the controller with the primary system which is given in equation (19)

$$C_P(s) = \frac{G_{ref}(s)}{G_{pri}(s)[1+G_{ref}(s)]} \quad (18)$$

$$C_P(s) = \frac{2.25s^5+13.725s^4+35.21255s^3+48.375s^2+34.7625s+10.125}{8s^6+39.6s^5+116.48s^4+188.76s^3+188.37s^2+89.1s+20.25} \quad (19)$$

Now let us calculate the transfer function of the controller with the reduced order system which is given in equation (21)

$$C_c(s) = \frac{G_{ref}(s)}{G_{comp}(s)[1+G_{ref}(s)]} \quad (20)$$

$$C_c(s) = \frac{2.25s^4+14.416s^3+43.704s^2+60.218s^1+41.153}{8s^5+41.729s^4+123.422s^3+204.12s^2+196.23s+82.307} \quad (21)$$

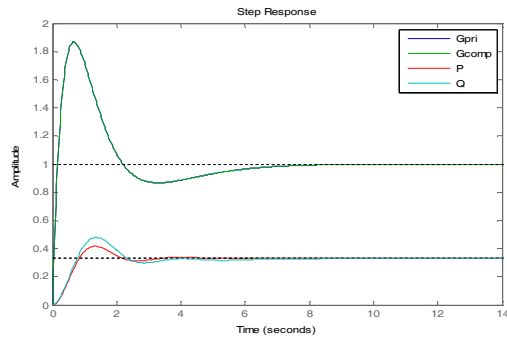The closed loop transfer function when the controller is attached to the primary system; is given below:

$$P(s) = \frac{C_P(s)\,Gpri(s)}{[1+C_P(s)Gpri(s)]} \quad (22)$$

$$P(s) = \frac{18s^7+123.3s^6+368.554s^5+625.728s^4+638.78s^3+386.328s^2+130.276s+20.25}{8s^9+71.6s^8+332.88s^7+991.98s^6+1973.564s^5+2645.068s^4+2334.8s^3+1289.568s^2+409.726s+60.75} \quad (23)$$

The closed loop transfer function when the reduced order controller is attached to the primary system; is given below:

$$Q(s) = \frac{C_c(s)\,Gpri(s)}{[1+C_c(s)Gpri(s)]} \quad (24)$$

$$Q(s) = \frac{18s^6+128.828s^5+440.628s^4+772.8s^3+777.94s^2+367.354s+82.306}{8s^8+73.729s^7+348.338s^6+1051.281s^5+2153.906s^4+2907.471s^3+2496.558s^2+1171.349s+246.92} \quad (25)$$

Fig. 3: I. Step responses of Gpri(s) and Gcomp(s)
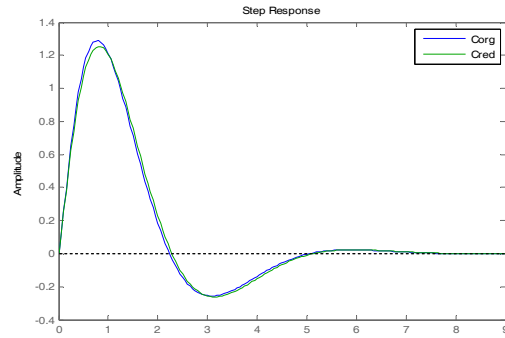**II. Step responses of P(s) and Q(s)**

8.   NUMERICAL EXAMPLE

(Chebyshev-Pole Clustering Approximation)

Consider a fourth-order system described by the transfer function as:

$$G_{org}(s) = \frac{s^3 + 7s^2 + 24s + 24}{s^4 + 10s^3 + 35s^2 + 50s + 24} \quad (6.1)$$

Choose a reference model which satisfies the control specification. In this example, a standard second-order transfer function is chosen with damping ratio $\epsilon = 0.7$ and natural frequency $w_n = 1.5$ rad/sec.

$$G_{des}(s) = \frac{2.25}{s^2 + 2.1s + 2.25} \quad (7.1)$$

Thus, transfer function G_{ored} (s) becomes by using Chebyshev-Pole Clustering Approximation :

$$Gred\,(s) = \frac{6.397 + 1.06s}{s^2 + 7.99s + 8.8885}$$
$$(8.1)$$

$$G_{frr}(s) = \frac{2.25s^3 + 58.77s^2 + 394.9506s + 316.21095}{s^5 + 28.22s^4 + 232.6356s^3 + 567.92876s^2 + 690.08082s + 316.21095}$$
$$(9.1)$$

$$G_{fro}(s) = \frac{2.5\,s^5 + 63.135s^4 + 586.552s^3 + 2490.485s^2 + 4299.075s + 2263.14}{2s^7 + 30.1s^6 + 339.256s^5 + 1802.923s^4 + 4824.57s^3 + 7509.16s^2 + 6411.67s + 2263.14}$$
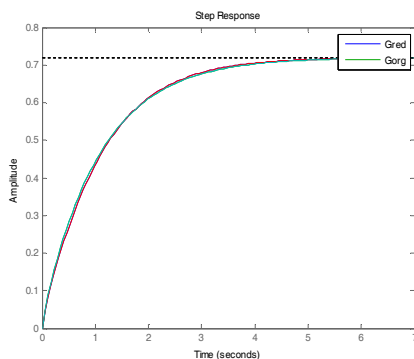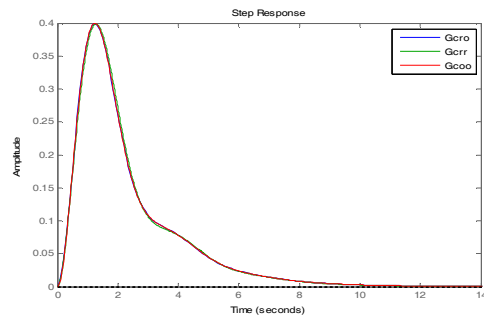$$(10.1)$$

Fig. 4.

Fig. 5.

Fig. 6.

9.   CONCLUSION

The first method for controller design is based on Pole Clustering and Chebyshev Polynomial technique method pertaining to unit step input has been developed. The original higher order plant is first approximated by a low order model using chebyshev-pole clustering technique and a controller is designed for this low order model. A reference model G(s) with assuming damping ratio, $\epsilon = 0.7$ and natural frequency, $w_n = 1.5$ radian/sec has been chosen. In the second method a controller is designed using the classical approach. The system is first approximated by a low order model using Pade approximation technique and a controller is designed for this low order model. For the design of the controller, a reference model $G_{ref}(s)$ with has been chosen. In both the methods reduced order controller is attached to the original higher order system and it was found that the step response of the primary system with reduced order controller is a good approximant to the step response of primary system with higher order controller. The present technique has been applied to the continuous systems further it would be interesting to implement the same idea to the discrete systems as well.

10.   REFERENCES

[1] M. F. Hutton and B. Friedland, "Routh approximationsfor reducing order of linear time-invariant systems," *IEEE Trans. Automat. Contr.*,vol. AC-20, pp. 329–337, 1975.

[2] Y. Shamash, "Stable biased reduced-order models using the Routh method of reduction," *Int. J. Syst. Sci.*, vol. 11, pp. 641–654, 1980.

[3] T. C. Chen, C. Y. Chang, and K. W. Han, "Stable reduced - order Padé approximants using stability equation method," *Electron. Lett.*, vol. 16, pp. 345–346, 1980.

[4] J. Pal, "Improved Padé approximants using stability equation method," *Electron. Lett.*, vol. 19, pp. 426–427, 1983.

[5] N. Ashoor and V. Singh, "A note on low order modeling, " *IEEE Trans. Automat. Contr* ., vol. AC-27, pp. 1124–1126, 1982.

[6] V. Singh, "Nonuniqueness of model reduction using the Routh approach,"*IEEE Trans. Automat. Contr.*, vol. AC-24, pp. 650–651, 1979.

[7] "Remarks on system simplification using the Routh array," *Proc. IEEE*, vol. 69, p. 662, 1981.

[8] R. Prasad, "Padé type model order reduction for multivariable systems using Routh approximation," *Comp. Elec. Eng.*, vol. 26, pp. 445–459,2000.

[9] B. Bandyopadhyay, O. Ismail, and R. Gorez, "Routh-Padé approximation for interval systems," *IEEE Trans. Automat. Contr.*, vol. 39, pp.2454–2456, 1994.

[10] B. Bandyopadhyay, A. Upadhye, and O. Ismail, "γ - δ Routh approximation for interval systems," *IEEE Trans. Automat. Contr.*, vol. 42, pp.1127–1130, 1997.

[11] C. Hwang and S. F. Yang, "Commentson the computation of interval approximants," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 1782–1787,

[12] G. V. K. R. Sastry, G. R. Rao, and P. M. Rao, "Large scale interval system modeling using Routh approximants," *Electron. Lett.*, vol. 36,pp. 768–769, 2000.

[13] Vimal Singh and Dinesh Chandra,"Improved Routh-Pade approximants: A Computer-Aided Approach, *IEEE 2004.*

[14] M. Aoki, "Control of large-scale dynamic systems by aggregation,"*IEEE Trans. Automat. Contr.*, vol. 13, pp. 246–253, 1968.

[15] N. K. Sinha and B. Kuszta, *Modeling and Identification of Dynamic Systems*. New York: van Nostrand Reinhold, 1983, pp. 133–163, ch. 8.

[16] Y. Shamash, "Order reduction of linear systems by Pad´e approximation methods," Ph.D. dissertation, Univ. of London, 1973.

[17] M. F. Hutton and B. Friedland, "Routh approximation for reducing order of linear time invariant system," *IEEE Trans. Automat. Contr.*, vol. 20,pp. 329–337, 1975.

[18] K. Glover, "All optimal Hankel-norm approximation of linear multivariable system and their L1 error bounds," *Int. J. Contr.*, vol. 39, pp. 1115–1193, 1984

[19] Y. Shamash, "Model reduction using Routh stability criterion and the Pad´e approximation," *Int. J. Contr.*, vol. 21, pp. 475–484, 1975.

[20] "Linear system reduction using Pad´e approximation to allow retention of dominant modes," *Int. J. Contr.*, vol. 21, pp. 257–272, 1975.

[21] B. Bandyopadhyay and S. S. Lamba, "Time domain Pad´e approximation and model-Pad´e method for multivariable systems," *IEEE Trans.Circuits Syst.*, vol. 34, pp. 91–94, 1987.

[22] B. Bandyopadhyay, A. Rao, and H. Singh, "On Pad´e approximation for multivariable systems," *IEEE Trans. Circuits Syst.*, vol. 36, pp. 638–639,1989.

[23] B. D. 0. Anderson and Y. Liu,"Controller reduction: Concepts and approaches," in ***Proc. Amer. Conrr. Conf,*** Minneapolis. MN, June 1987, pp. 1-9.

[24] D. Gangsaas, K. R. Bruce, J. D. Blight, and **U.** -L. Ly, "Application of modern synthesis to aircraft control: Three case studies," ***IEEE Trans. Automat. Contr.,*** vol. AC-31, pp. 995-1104, Nov. 1986.

❑❑❑

OPTIMIZATION AND AUTOMATION OF P2P PROCESS LEADS TO PROFITABILITY TO THE INDUSTRY

Suresh Kumar
Research Scholar, Head Supply Chain,
Bharat Electronics Limited(BEL), Ghaziabad, Uttar Pradesh, INDIA.
Corresponding Email ID: iimm4delhi@gmail.com

ABSTRACT

Procure to pay process are at the core of the company's operations but fraught with risk, uncertainty and inefficiency. Howeverits impacts on the business is tremendous.If we look back on the traditional business that is very hectic and cumbersome. The user department send the requirements to material control departments than the materials department consolidate the all purchase requisition segregate them and send to purchase department than purchase department start its function of enquires, quotations, tendering process and select the supplier out of the many supplier than place the order on the selected one than supplier takes his own time and deliver the goods to the buyer premises as per the terms and conditions which are stipulated in the purchase order. The goods delivered by the supplier received at the premises of the buyer, where these are counted and mach with the invoice/delivery challan and goods receipt note is prepared and offered for inspection to quality department to check the materials as per the specification mentioned in the purchase order,if the materials is found Ok as per the requirement specified in the PO than goods received note is forwarded to finance department dully signed by the inspection authority.Afterwards finance department look in to the invoice of the suppler from all angles with the terms and condition of the PO and finally make the payment.If materials is rejected than supplier is informed and advised to pick-up the materials with in 15 days on receiving the letter from customer end.These all activities takes months together. There is no transparency in the system,payment is not made on time to the supplier after getting right materials on time.There is lot of ambiguity in each steps in the system, employees are adamant, reactive and back office functionary.Semiautomatic technology is applied to work with the system,a legacy software is used in in the whole process.Supplier is not considered a partner.Customer and supplier do not have a long terms relations.There is lack of standardization at both side supplier as well as customer.Lead time is uncertain,excess inventories is compiled at various level due to un- certainty of receiving the needed materials on time.Now we have to automateandoptimizewith unprecedented technologythe P2P process which is the gate way to the industries growth.

Keywords : Unprecedented, Harnesh, Ubequitous, Ambiguity, etc.

1.   INTRODUCTION

Procure to Pay is one of the Enterprise Resource Planning (ERP) technique that can be acquired by small or large-scale organizations, for better management and less dependencies in the system it is the end to end journey of acquiring materials needed to effectively run a business, while facilitating the complete bill payment steps for the supplier.

Procure to Pay is developed from the concept of internal control system (ICS), where the work of one person is checked by another person to minimize the errors, mistakes and to record transactions as transparent as possible.
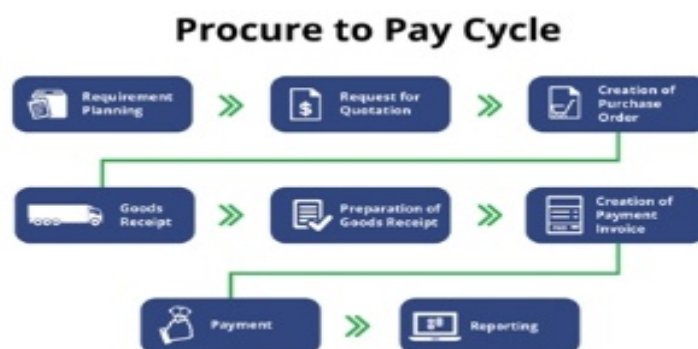

Fig.1 : Procure to Pay Cycle

P2P is cycle of steps that has the following parts:

- **Requisition -** When the inventory or the product needs to be sold to the customer a requisition is raised to purchase department, the purchase department will evaluate the requisition form and analyze the quantity, item and description of the product that they need.
- **Request for Quotation** - Then they check and identify different suppliers or vendors available in their region and ask for their quotation.
- **Quotation Analysis** - Now the purchase department will do the quotation analysis which is checking whether the terms and condition, quantity and quality are matched by the suppliers.
- **Selection of Supplier** - Based on the quotation analysis and many other factors like accountability and past records of the vendors a decision is taken.
- **Raising Purchase Order(PO)**- Then a PO is raised which is issued to the selected supplier or vendor, When Purchase department is issuing this PO this becomes the last set of agreement between the supplier and purchase department of the company in which the department mentions the essential details for the proper delivery of the goods for ex- dates of delivery, final amount, quantity, terms and conditions etc. The vendor is legally required to supply the goods according to those conditions to their respective locations.

- **Receipt of Goods/Materials** - Once the delivery is made it is checked at the place of reception, it might be a warehouse or a manufacturing factory, the receiver checks the quantity physically and matches everything with the Invoice/Delivery challan/PO, then they prepare a good's receipt note for the carried out Inspection and forward to finance Department.
- **Submitting documents to Finance** - Warehouse team puts together all the documents relevant to the purchase of the good to give it to finance department, supplier directly sends the invoice to the finance team and warehouse need not be aware of the finance portion for less coupling of the jobs.
- **Verifying the invoice with PO/Receipt**- The finance team will match the goods and material receipt generated by warehouse to the invoice sent by the supplier.
- **Final Payment to Supplier**-The final step in the procure-to-pay process after approving the purchase order for payment is sending the
- payment, and entering the payment into accounting system.

Below, we can see the impact that digitization has had on entire industries and businesses. Travel, retail, media, financial services have all in some way been significantly improved:
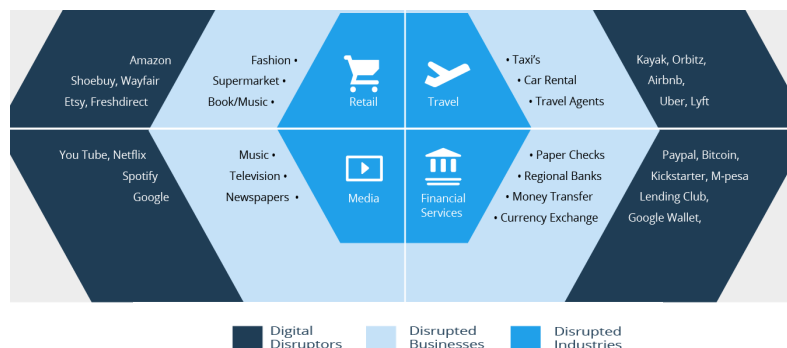

Fig. 2

2. PROCURE TO PAY HAS CONSIDERABLE IMPACT ON THE BUSINESS.

There are many checks and balances are put in the process and authorization of many managers are needed at many places. Since the process is spread across so many departments.

There are many organization/company who conduct their business manually and using excessive paper work hence facing risk of documentation errors because of that always delay in processing.In many companies you may see lack of communication and cohesiveness among various divisions. And even between the personnel working in the same unit.

There could be unwanted delays in the documentation travelling across the different departments, which might harm the buyer supplier relations.As survey conducted by many IT companies in the past they ranked the P2P process highest among business process most vulnerable to fraud ,errors and time consuming.

3. UNDETECTED FINANCIAL LEAKAGE

Companies do not realize the efficiencies that may be gained through digitization in the business process for examples invoices, payments reviewed through manual authorization/approvals, this process is not only cumbersome and time consuming but also fail to take advantage of early payment discount. Subsequent to this

the manual approval process leaves the door open to big fraud through post approval modification,

4.   MAINTENANCE OF VENDORS RECORD.

In large company typically have multiple employees with access to the vendor master and often there are duplicate entries of suppliers in the systems, without tight control on the vendors data the chances of fraud ,financial leakages and mismanagement of funds remains significant with out a proper digitized P2P solution.

Financial loss between Purchasing-Receiving-Accounts payables. As soon as the materials reaches to the inwards goods stores to make the suppliers payments ,manuals matchings for each and every transaction is time consuming as well as error prone. Only solution to this problem is atomization of P2P process, automated machine consolidate the matching rules of different departments and streamline and accelerate the process and yield satisfactory results.

Process automation and software benefits: By uniting these systems, businesses can become more efficient, save time and money, and ultimately provide better shareholder value and an improved customer experience, it's all about getting rid of paper and getting more efficient functioning of the company

Procure to pay harnessed with unprecedented technology helps reducing errors and improve efficiency by linking the entire process under one system. Procure to pay aims to get the right items into the hands of the right people when they need them.

There are departments like, Supply chain,, inventory, production, accounting, and finance team members are involved in purchase to pay systems each bringing unique expertise and contributing towards better results for the company. But with such a diverse group involved, team members face many opportunities for inefficiencies and slowdowns. Moreover, depending upon the size of the organization and areas covered by the product the location of each of these different departments may vary. Due to such large differences in the locations there can be communication lags and coordination issues, but the automated P2P software makes the entire process as if happening under the same roof, the software maintains a centralized database for the entire organization which helps the entire system to work as one.

The P2P process is critical to every organization. It is imperative that the process should be efficient and effective as possible. The most basic aspects would be automating transactions, managing the flow of information and routing approvals. Today many companies are not taking full advantages of automating the procure to pay process, which could help tighten control lower resources needs for manual validations and processing data inputs, which in turns lowers the risk of errors that are inevitable with high volumes of manuals inputs.

RFID technology to be Introduced: This technology has risen to become a revolutionary element in supply chain management. RFID ensures that the right goods are available in the right place with no discrepancies and zero errors. There are the following ways RFID technology will provide benefit to P2P process.
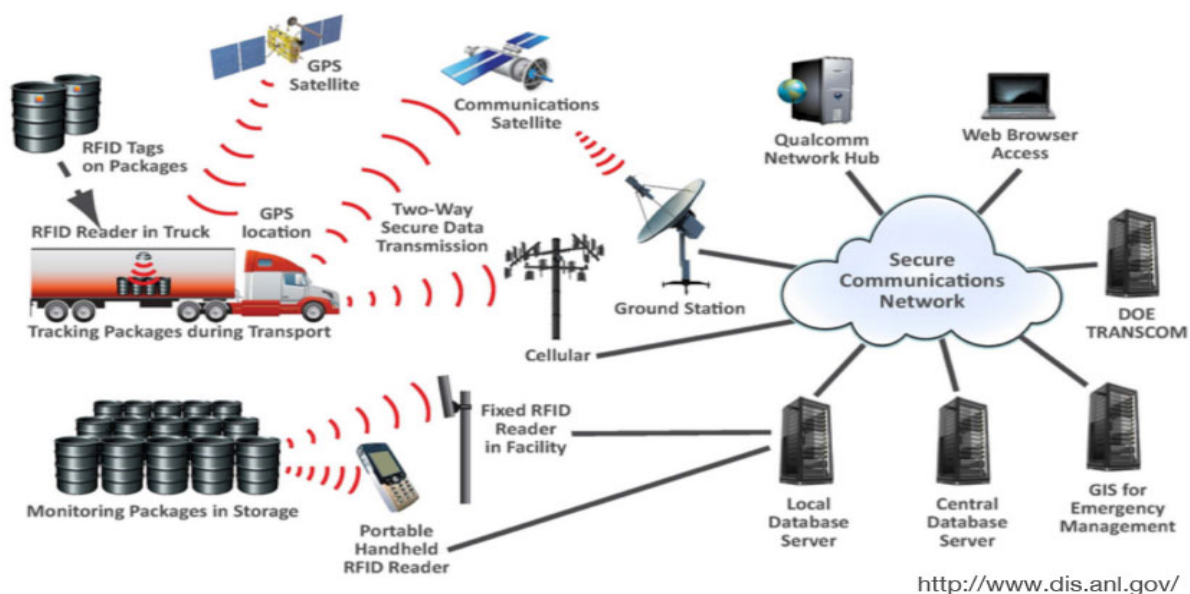

Fig. 3: P2P Process

Speeding the receiving process- RFID tagged product can be scanned and directly/automatically inserted into

inventory mesh.Hence drastically saving man -hours.

Inventory status-how much is the inventory levels,how much needed in the future may be found out easily through RFID display monitor,how much current stock is available in the location and how much sold out all type of current information is one click from away from hand scanner and also eliminate the need for manual reconciliation.

TRACK AND TRACE -For compliance purposes also for perishable items and pharmaceuticals and hazardous materials which are RFID tagged may give real time tracking and tracing of location and establish a chain of custody. Which may fulfill the statutory requirement of rules and regulation of govt.

ASSET MANAGEMENT- assets with RFID tagged gives real-time information as regard to asset location hence avoid compilation of excess inventories in the warehouse.

THEFT DETERRENT- The high valued items kept on location with RFID tagged the real time notification provided by them when themes leave the secure areas hence reducing the likelihood of inventories shrinking due to theft/pilferage. There for RFID technology which are excessively used in western countries now it is being used in retail stores in our countries also.This technology will become truly ubiquitous as the major retailers and govt. agencies continues to push suppliers of its adoption.It is transforming the P2P process by speeding the process from somany angles.

5. CONCLUSION

Procure to pay processyields benefit after automation and optimization. The automation of P2P service is improving collaboration between buyers and suppliers and suppliers are considered the business partner he is one of the stake holder of the company and homogenous approach is generating more supply chain value. Digital enablement of P2P process helps buyers and suppliers to come more closer and have better business relation After implementing collaborative cloud based P2P platform it is easier to put standard business rules and information /workflow in place which offering a shared identity to do every one a harmonious business. Automation helps buyer to interact with all the related stake holder fully and, quickly.by optimizing P2P it is possible to keep accurate information as regard to order,price,deliveryetc.Reduces manualprocessing, removes manual control, improve data quality and streamlines document matching. Data flows freely between supplier's end to buyer are set right After transforming the P2P process every thing is set right. Payments to supplier is automatically made, as soon as inspection is over at buyer's end. There is no ambiguity seen in the whole process. Companies are getting tangible and intangible benefit ultimately all round growth may be seen in the corporates.

6. REFERENCES

[1] Celtrino Marketing
[2] Google
[3] World Economic Forum
[4] Ivalua website
[5] MMR A journal of IIMM

❑❑❑

DETAILED STUDY OF INFORMATION TECHNOLOGY INFRASTRUCTURE LIBRARY SERVICE LIFECYCLE

¹Shruti Garg* and ²Dr. Anuranjan Misra
¹Research Scholar, CSE Department, Centre for Research Studies, NIU, G.Noida, Uttar Pradesh, INDIA.
²Prof. and Director, Accurate Institute of Management & Technology, G.Noida, Uttar Pradesh, INDIA
*Corresponding Author Email ID: shruti1707@gmail.com

ABSTRACT

The service sector is growing day by day. The more the products are launched in the market, the more they require service. For fulfilling the needs of the business, managing the services is a first priority of the IT organizations. IT Service Management is an increasingly important area of study for all IT professionals. IT Managers are realizing that whatever the technology in use, the requirements to manage it efficiently and effectively and to deliver services that are aligned to the business requirements have never been more important. 'Information Technology Infrastructure Library (ITIL) is a set of best practices for IT service management (ITSM) that focuses on aligning IT services with the needs of business.' [1]. ITIL is a service framework, on which Service Management depends. ITIL makes sure that IT services are aligned with the needs of the business and provides support for the services offered by them. It offers best practices for the organizations to transform IT and grow. In this paper, we shall study the lifecycle of ITIL framework. We shall also try to bring few changes which could help organizations in meeting customer expectations.

Keywords: ITIL, ITSM, Service Strategy, Service Design, Service Transition, Service Operations, CSI.

1.   INTRODUCTION

Sometimes, the organizations seek business to own and operate assets, which could be an overhead for the business, in terms of resource and time management. To prevent this, organizations started seeking alternatives from the third party vendors. These third party vendors own the infrastructure and provide services. Service is basically a combination of People, Processes and Technology, which is used to provide support to the customer's business processes. As per the ITIL definition, 'service is a means of delivering value to customers by facilitating outcomes customers want to achieve without the ownership of specific costs and risks.' [1]

As per the definition of ITIL provided by the Wikipedia, 'Information Technology Infrastructure Library (ITIL) is a set of best practices for IT service management (ITSM) that focuses on aligning IT services with the needs of business.' [2] ITSM employs the ITIL processes to provide better services to the users and focusses on continuous improvement Telecommunications Agency, commonly called as CCTA, created ITIL in the 1980s, which made V1 as the first version of ITIL containing more than 30 volumes. The number was reduced and made up into logical sets of 8 in the V2 version of ITIL. As per the Wikipedia, 'the organization issued V3 in 2007, which

consisted of 5 core processes, namely, Service Strategy (SS), Service Design (SD), Service Transition (ST), Service Operations (SO), and Continual Service Improvement (CSI).' [2] Later, in July 2011, the updated version of V3 was released. Fig. 1. represents the 5 publications of ITIL Core.
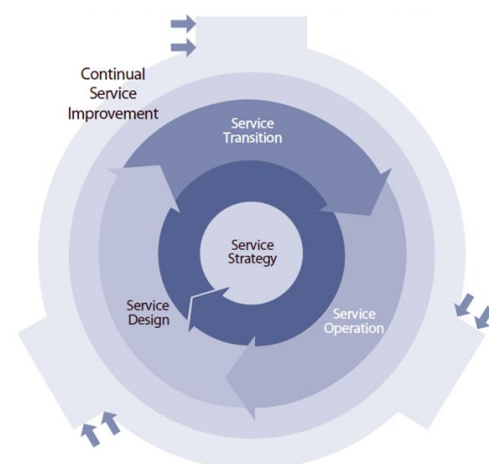

Fig. 1: ITIL Core [5]

As per the dictionary definition, 'strategy is a plan of action designed to achieve a long-term or overall aim.'

[3] Service Strategy is the first stage of the ITIL Service Lifecycle. 'It asks *why* something is to be done before thinking of *how*.' [4] In layman terms, this stage is the planning stage for the service providers. If wrong strategy or plan is adopted, the entire service will go wrong. For example, if the needs of the customers are not clear, the companies will face challenges in selling the product. For example, the CRT televisions are completely replaced by the LCD televisions and now the industry is more focused on the next better version of the product. The focus of the industry is to make advancements to the products by using better technologies, keeping in view the business needs.

Service Strategy defines the perspective, position, plans and patterns for the service provider, which would help them in meeting the organization's business results. Perspective defines the vision and mission for the service provider. Position helps in finding out how the service provider should compete with other service providers in the market. Plans provide the insight into how the service provider should take every step towards the desired goal. Lastly, patterns help the service provider in finding out the repeatable actions to meet its objectives.

Service Design phase is the next phase in the service lifecycle. Various processes under Service Strategy module are: 1) Strategy Management for IT Service, which ensures that the strategy is achieving its business outcomes. 2) Demand Management, which understands the demand of the customer, in terms of the services, and the finds the capability to fulfil them.3) Service Portfolio Management, which ensures that the appropriate mix of services are provided to the customer to meet its requirements. 4) Financial Management, which creates a proper balance between the cost and the quality of the service. 5) Business Relationship Management, which ensures healthy relations are maintained between the client and the service provider. Also, it ensures that the service provider understands the changing needs of the business over time.

2. SERVICE DESIGN

Service Design phase is the next phase in the service lifecycle. Once the strategy is in place, the design, to implement the ideas, becomes the focus of the service lifecycle. Strategy tries to answer the *why* of the service requirement; design tries to answer the *how* of it. Design is always planned keeping in mind the business objectives. Design covers not only the new service, it also covers the modification in an existing service. 'Service Design identifies service requirements and devises new service offerings as well as changes and improvements to existing ones.' [6] The objectives of Service Design phase are: 1) Ensuring that the service is adaptable to the agreed future requirements of the business. 2) Identifying and managing the risks, so that these can be removed before the service is transitioned

for the customers. 3) Ensuring that the service is transitioning as per the requirements of the customer. 4) Considering how the effectiveness of the service would be measured.

The complete list of processes under Service Design phase is: 1) Design Coordination, which maintains the coordination between the people and technology. 2) Service Catalogue Management, which maintains a single source of all the services and information related to these services. 3) Service Level Management, which ensures that the service is being provided to the customer at the agreed levels. For this purpose, Service Level Agreements are made between the customer and the service provider. 4) Availability Management, which ensures that the service is available to be performed at the agreed service time. There should be minimum outage or planned interruption to the service. 5) Capacity Management, which ensures that the capacity is maintained at a level so that the service is not impacted. 6) IT Service Continuity Management, which uses the methods of the risk assessment to reduce risks to the services. 7) Information Security Management, which ensures the confidentiality, integrity, and availability of the organization's data and services. 8) Supplier Management, which describes the best practices for the service providers to manage the suppliers, in order to get the work from the suppliers, which helps the service providers in providing up to the mark service to the customers.

3. SERVICE TRANSITION

Service Transition is the next phase in the service lifecycle. Whatever plans are devised and whatever agreements are made with the client, all are implemented in this phase of the lifecycle. The deployment/implementation is done under the release cycles. Management should be aware of the release planning and design, for their proper involvement.

The complete list of processes under Service Transition is: 1) Transition Planning and Support, whose 'objective is to ensure that the purpose and goals are achieved in a consistent and repeatable manner through the production of comprehensive plans that ensure effective and efficient transition into the live environment.' [4] 2) Change Management, which ensures that the changes within the organization are implemented without disruption to the services. It not only rejects changes, which might appear a risk for the organization, but also encourages some of these if potential for new business is realized. 3) Service Asset and Configuration Management, which ensures that the assets, which are required to deliver the service, are being managed properly. It also shows the relationship between the assets and their configuration. These assets are referred to as Configuration Items (CI). 4) Release and Deployment Management, which manages the release packages and ensure that the release items are properly planned and are deployed in a timely manner.

5) Service Validation and Testing, which ensures that the deployed changes meet the specified design and the needs of the business. It should be delivered with the agreed levels of warranty. 6) Change Evaluation, whose purpose is to evaluate the changes that are going to implement during the deployment, to ensure that the changes have no impact on the business and the services. 7) Knowledge Management, which keeps a record of information and knowledge in a single database for the ease of access to the users.

4.   SERVICE OPERATIONS

Service Operations is the next phase in the service lifecycle. Once the service is transitioned for the customers, it must be supported in order to provide better experience to the customer.The purpose of Service Operations is to deliver and manage services to the customers as per the agreed levels. In this phase only the customer experiences the services. In case of any issues with the provided service, users contact the service provider through Service Desk. This phase also controls the access to the services.

The complete list of processes under Service Operations is: Event Management, whose objective is to detect the events and their cause. It identifies issues which occur with a Configuration Item (CI)and create an incident if issues found. 2) Incident Management, which identifies and manages the issues reported in a service, in the form of incidents. Incident is an unplanned interruption in the service. Incident Management helps in identifying the issues but the root cause is identified under Problem Management. 3) Problem Management, which identifies the root cause of the incidents which had significant impact on the number of users. Purpose of Problem Management is to document, investigate and provide permanent fix to the incidents reported. 4) Request Fulfilment, which is responsible for managing the service requests from the users. Service Request could be a request for access to an application, or it could be a request for password reset etc. 5) Access Management, which is responsible for creating and managing people accounts, who need to use the provided service. This process deals with security of the application or service.

5.   CONTINUAL   SERVICE IMPROVEMENT CONTINUAL

Service Improvement is the last process of the service lifecycle. But this phase is functional in all the phases of the lifecycle. The main objective of Continual Service Improvement is to ensure that the IT services are aligned to the changing business needs. It ensures that improvements are identified and implemented for every phase of the lifecycle. 'This process reviews, identifies and recommend improvement opportunities in each of the lifecycle stage.' [4]

6.   CONCLUSION

'The ITIL framework provides a robust framework for identifying, planning, delivering and supporting IT Services that can be adapted and applied to all business and organizational environments.' [7] It manages the business risk and service disruption or even failure. It improves the service delivery and helps maintain healthy relationships between the service providers and the customer.

REFERENCES
[1]   OGC, "ITIL Version 3 Service Strategy".
[2]   ITIL, URL: https://en.wikipedia.org/wiki/ITIL, web link accessed on 11/28/2016.
[3]   Strategy,   URL: https://en.oxforddictionaries.com/definition/strategy, web link accessed on 02/19/2018.
[4]   Helen Morris and Liz Gallacher, "ITL Intermediate Certification Companion Study Guide".
[5]   OGC, "ITIL Version 3 Service Design".
[6]   ITIL   Service   Design,   URL: https://wiki.en.it-processmaps.com/index.php/ITIL_Service_Design,   web link accessed on 02/25/2018.
[7]   Key   Benefits   of   ITIL,   URL: https://www.axelos.com/best-practice-solutions/itil/key-benefits-of-itil, web link accessed on 02/26/2018.

❑❑❑

ROUTING PROTOCOLS FOR MANET

[1]Ajay Kumar Sharma* , [2]Dr. D.K Chauhan and [3]Dr. Anuranjan Mishra
[1]Research Scholar, CSE Department, [2]Professor & Director CRS,
[1,2]Noida International University, G.Noida, Uttar Pradesh, INDIA.
[3]Prof. & Director, Accurate Institute of Management and Technology, G.Noida, Uttar Pradesh, INDIA.
*Corresponding Author Email ID: hiajai@yahoo.com

ABSTRACT

Ad Hoc network is more convenient and cheaper than the networks with infrastructure in the usage and setup. As to wire network, the router and the terminals are also existed in Ad Hoc network. Compared with the roles of nodes in the wire network, the major differences are which act two different roles meanwhile in Ad Hoc network. In practice, it is not only to study in communication security but also to setup the correct route becomes a very important subject. In this paper, we studied several papers[7].

Emerging types of Wireless Networks are Cellular Networks, Mobile Ad Hoc Network, wireless Sensor Networks, Vehicle Ad - hoc Networks, and Wireless Mesh Networks. With the advance of wireless communication technologies, small - size and high - performance computing and communication devices like commercial laptops and personal digital assistants are increasingly used in daily life. After the success of second generation mobile system, more interest in wireless communication was started.[3] This interest has led to two types of wireless networks: infrastructure wireless network and infrastructure less wireless network it is also called Mobile Ad - Hoc Network (MANET).The Infrastructure less wireless network consists of a network with fixed wired base stations and mobile nodes.

In Mobile Ad Hoc Network all nodes are mobile; they connected dynamically in an arbitrary manner and each node behaves as a router and it takes part in discovery and maintenance of routes to other nodes in the network. The network structure changes dynamically due to node mobility.[1] Essentially Mobile Ad Hoc networks are suitable when infrastructure is not present or difficult or costly to setup or when network setup is to be done quickly within a short Period. They are very attractive for tactical communication in the military and rescue missions (Natural disasters like earthquake, flood, tsunami, and fire). One best real life example is attack on world trade center at New York in USA in 2001. They are also expected to play an important role in civilian for a such as convention centers, conferences, and electronic classrooms . This research paper surveys the three different areas of research fields, routing protocols, simulation tools and mobility models, these fields are interrelated to each other. The previous research papers have not surveyed these three areas of research fields combinedly.[4] Hence this paper is very useful to research community to propose and develop new routing protocols using simulator in Mobile Ad Hoc Networks.

Keywords : MANET, Routing Protocol, DSR, DSDV.

1.   INTRODUCTION

Today, we see two kinds of wireless networks but the difference between them is not as obvious as it seems. The first kind and most used today is a wireless network built on top of a "wired" network and thus creates a reliable infrastructure wireless network. The wireless nodes are able to act as bridges in a wired network is shown in fig.1. This kind of wireless nodes are called base - stations. An example of this wireless network is the cellular - phone networks where a phone connects to the base - station with the best signal quality. [2]When the phone moves out of range of a base - station, it does a "hand - off" and switches to a new base - station within reach. The "hand - off" should be fast enough to be seamless for the user of the network. The second approach, called Ad hoc, does not rely on any stationary infrastructure. All nodes in ad hoc networks are mobile and can be connected dynamically in an arbitrary manner. Each node in such networks behaves as a router and takes part in discovery and maintenance of routes to other nodes. In the MANET every node can perform the role of host as well as router, thus nodes, which are out of transmission range, can be accessed by routing through the intermediate nodes.

A. EVOLUTION OF MANET
The Evolution of Mobile Ad Hoc Networks as follows:
- In 1970, Normanl Abramson and his fellow researchers at the University of Hawaii invented ALOHA net.
- In 1972 DODARPA (Department of Defense Advance Research Project Agency) Packet Radio Network

- In 1980 Survivable Radio Networks (SURAN).
- During 1980 emergence of Internet Emerging Task Force (IETF), termed the mobile ad hoc networking group.
- In 1994 emergence of Bluetooth by Ericsson.

B. CHARACTERISTICS OF MOBILE AD HOC NETWORKS

The main characteristics of the MANETs as follows:

- Dynamic Topology,
- Bandwidth Constrained,
- Variable Capacity Links, Energy and Bandwidth Constrained Operation
- Limited Physical Security
- Quickly Deployable

C. APPLICATIONS OF MOBILE AD HOC NETWORKS

The following are some well – known applications of MANET are described given below

- Military
  - o Automated battlefield
  - o Special operations
  - o Homeland defense
- Civilian
  - o Disaster Recovery (flood, fire, earthquakes etc)
  - o Law enforcement (crowd control)
  - o Search and rescue in remote areas
  - o Environment monitoring (sensors)
  - o Space/planet exploration
- Commercial
  - o Sport events, festivals, conventions
  - o Patient monitoring
  - o Ad hoc collaborative computing (Bluetooth)
  - o Sensors on cars (car navigation safety)
  - o Car to car communications
  - o Networked video games at amusement parks, etc

D. RESEARCH CHALLENGES

Mobile Ad Hoc Networks (MANETs) has become one of the most prevalent areas of research in the recent years because of the challenges it pose to the related protocols.[5] The MANET is the new emerging technology, which enables users to communicate without any physical infrastructure regardless of their geographical location the following are research areas in the MANET.

- Reliability
- Routing
- Scalability
- Security
- Internetworking
- Multimedia
- Energy consumption optimization
- Quality Of Services
- Multipath Routing
- Cross layer design
- Unicast, Multicast, Broadcast, Geocast routing

protocols design

- MAC protocol design, Network layer Transport layer and application layer design
- Next generation hydride network protocols

2. ROUTING IN MOBILE AD HOC NETWORKS

MANETs are considered as autonomous, self - configured, multi - hop wireless networks. They don't rely on any stationary infrastructure; no centralized control and they are quickly deployable anywhere at any time. In the MANET, all nodes are mobile that cooperate in friendly manner, they are connected dynamically and they have the ability to leave existing nodes from the network and enter new nodes into the network at any time, due to this, the topology of the MANET is highly dynamic and it frequently changes. In the MANET, every node the route discovery and route maintenance phase. The MANEs are widely used in military, civilian and Commercial applications. The routing is the most active research field in the MANET.[6] The routing protocols, which are developed for wired networks are not suitable for wireless networks, due to the mobility. The routing problem becomes more serious in MANET, due to following characteristics of MANET:

- Keeping track of residual battery power.
- Tracking of already used paths.
- Keeping back-up paths.
- Keeping track of hanging messages.
- On demand calculation and updates of routing tables.
- Sending the data packets at lower energy.
- Requiring a node to send packets with energy symmetrical to the distance rather than with fixed energy.
- Using a hierarchical routing technique.
- Using directional antennas.
- Each node is behaved as the hosts and route so it is easy to send packets through this.

2.1 PROACTIVE PROTOCOL

The proactive protocol is also known as the "table driven routing protocol". As the name suggests that, the proactive or the table driven routing protocol, are properly maintained the routing information or packets even before it is needed.In proactive protocol, every node maintains one or more table representing the complete structure of topology of the network and the table under the protocols are regularly updated to maintain up-to-date routing information from each node to every other node. In other words routes packets or information is kept in the routing tables and is symmetrically updated as the network topology changes.[9] To maintain and to be always updated information, topology information needs to be exchanged between the nodes on a regular basis. On the same time, routes will always be available on the request.Some of the protocols which come under the table driven protocols are DSDV, WRP and OLSR.

2.2   REACTIVE PROTOCOL

The reactive protocol is also known as on-demand protocols. As its name shows that reactive or on-demand protocol do not maintain routing or routing activity or packets at the nodes of the network if there is no communication.

It's completely opposed to the proactive routing protocol. Basically reactive routing protocol is a bandwidth efficient on-demand routing protocol for these network. In this protocol if one node wants to send a packet or information to the another node then this protocol searches for a path or route in an on-demand manner that means it does not follow any specific path to send the packet to one node to another node.[10]

So through the reactive protocol it reduces the overhead encountered in the table driven protocols. This means that the routes are determined and maintained for the nodes that are on-demand to send a data or information to an exact destination. Some routing protocols comes under the reactive protocols are AODV, DSR and ABR.

2.3   HYBRID ROUTING PROTOCOL

Hybrid protocol is the collection of both the routing protocols or hybrid protocol is the combination of both reactive and proactive routing protocols. This comes to overhead the problems of proactive and reactive protocol.

Hybrid protocol is the well maintained combination of the proactive and reactive routing methods are used which are better than the both used separately.[8]

This protocol has the advantages of both table driven and on-demand protocol. The hybrid protocols are works on zone routing protocol (ZRP). ZRP was planned to reduce the problems of proactive routing protocols and discovery in reactive routing protocols also reduces the latency caused by route. It is well assumed that the most of the communication takes place between those nodes which are close to each other.Zone routing protocol provides framework to other protocols.

3.   METHODOLOGY

To resolve such issues many of researchers have produced so much ideas and innovations in this field. There are many reactive routing protocols based on AODV most of the work based on energy efficient routing, as energy efficiency routing or power is the major concern in ad-hoc wireless networks.

As each and every protocol has some advantages and some shortcomings, there is not any protocol which can perform better in every condition.[12] It depends upon the network parameters which decide the protocol to be used in specific condition:

I.   DESTINATION SEQUENCED DISTANCE VECTOR ROUTING (DSDV)

DSDV is based on table driven routing scheme for mobile ad-hoc network. It basically works to solve the routing loop problem. Each and every entry in the routing table has a sequence number, the sequence number would be positive if the links are maintained in the network else an odd number is used.

The number is generated by the destination. In DSDV each node maintains a table that contain the shortest distance and the first node on the shortest path to every other node in network.

II. DYNMIC SOURCE ROUTING (DSR)

It is the routing protocol for wireless mesh networks and a self-maintaining routing protocol for wireless network; it does not need any existing network infrastructure or administration.

DSR allow multiple routes to any destination and allows each sender to select and control the routes used in routing its packets.

III. ASSOCIATIVITY-BASED ROUTING (ABR)

ABR is a mobile routing protocol used for mobile or wireless ad-hoc network. The best thing about the ABR is that it is free from loops, deadlock and packet duplicates. As its name suggests that it maintains routes for only those sources that actually desire routes.

IV. AD-HOC ON DEMAND DISTANCE VECTOR ROUTING (AODV)

The ad-hoc on demand distance vector routing is one of the algorithm used in reactive routing protocol, basically it is a routing protocol designed for wireless and mobile ad-hoc networks. It is the improved version of destination sequenced distance vector (DSDV) routing in which all available wired network protocol can be useful to ad-hoc wireless networks with less modification. This AODV protocol minimizes the number of transmission by creating routes on-demand as opposed to all possible routes as in DSDV.

AODV protocol checks the entire routes table when source needs to broadcast data. This protocol makes routes to the destinations on demand and supports both unicast and multicast routing. So AODV is considered an on-demand algorithm and does not create any extra traffic for communication along links and these routes are maintained until or unless they are required by the source.[11]

In AODV the networks are quite until connection are established. In this process the AODV keeps track of all the nodes to remove traffic.

There are two main procedures in AODV:
- Route discovery
- Route maintenance

ROUTE DISCOVERY

It is simple to understand the route discovery starts when a

source node wants a route to destination to send data or packet.

First it check table if there is any current route to a destination or not to send data, if it founds a route, forward data packet to next hop node or else it starts route discovery process. Route discovery starts with the creation of a route request (RREQ) packet. That data packet contains the source nodes IP address, destination sequence; source nodes current sequence number and broadcast ID number.

ROUTE MAINTENANCE

This is done through route error (RERR) packet. The route is known as expired route if it is not use frequently. In the route maintenance each predecessor node and its adjacent node are active to maintain their job. A proper set of predecessor nodes is maintained for each routing table entry and indicating to its adjacent nodes to pass the data packet by using same path or route. If the data packet is passing through the route and in sudden the route breaks or the next hop link breaks these nodes are notified with rote error (RERR) packet. After that the route goes forward, each predecessor node in turn starts forward the RERR packet to its own set of predecessors, effectively erasing all routes using the broken link. Then this RERR packet transmits through the all failed link and then process is reinitiated if routes are still needed.

V. ZONE ROUTING PROTOCOL (ZRP)

Zone routing protocol is based on hybrid networking routing protocol which uses both proactive and reactive routing protocol to maintain the network and to send the data packet.

If a data packet is in the same zone as the origin, then it uses the already stored routing table to transmit the packet frequently. This process comes under the proactive protocol. On the other hand if the condition is opposite, i.e. if the route extended outside packet's initial zone, then the reactive protocol traces the each successive zone in the route to see whether the destination is inside that zone. And finally when the zone is confirmed as containing the destination node the stored route listing table is used to transmit the packet through the network.

Basically ZRP reduces the problem for longer routes that would be important if using proactive routing protocols throughout the entire route, while eliminating the delays for routing within a zone that would be by reactive routing protocol.

4. ENERGY EFFICIENT ROUTING ALGORITHMS FOR MANET'S

Energy efficiency is the major issue for mobile ad-hoc network and these energy efficiency algorithms are not just related to reduce or save the total energy consumption of the route but also increases the lifetime of each and every node in the network to maximize the lifetime of the network. The main goal of the energy efficient algorithm

is to maintain the functions of network as long as possible.

Green networking recently takes a most admirable place in the networking field, including a reflection on the in which networks are constructed and operated in such a way that not only cost and performance are taken into account but also their energy consumption and carbon footprint [1].

Applying green networking to wireless networks (WMNs) has rarely been described in the literature. Wireless mesh network is a new emerging technology which has been adopted as the wireless internetworking solution for the near future [2].

4.1 ENERGY CONSERVATION IN NETWORK LAYER

As network layer is responsible to create a self configured network and packet routing. To maintain the network technology network layer chooses the route and transmit the data to the most suitable node with which communication links forms.

Many of technique to the energy consumption for mobile ad-hoc network which minimizes the energy consumption without dismissing of the connectivity of the network.

Some approaches for the WMNs are as follows:
- The CDS (Connected dominating set)
- The SPAN
- The GAF (Geographic adaptive fidelity)
- EMM-DSR (Extended max-min dynamic source routing)
- The minimum energy routing
- Power aware routing
- The pulse
- The green clustering
- Cadet (Clustering and decision-tree-based)
- EAR (Energy aware routing)

4.2 ENERGY CONSERVATION IN DATA LINK LAYER

There are some approaches through which we can save the energy in data link layer:
- Power saving mode (PSM)
- Automatic power saving delivery (APSM)
- Power save multi poll (PSMP)
- Network allocation vector (NAV)
- Power aware multi-access protocol with signally (PAMAS)
- A sleep optimal fair attention schedulers (SOFA)
- Low energy adaptive clustering hierarchy (LEACH)
- S-MAC
- Virtualization
- Power efficient MAC protocol

5. CONCLUSION

The field of MANETs is rapidly growing and changing.

While there are still many challenges that need to be met, it is likely that such networks will see widespread use within the next few years. One of these challenges is security. Security of mobile adhoc networks has recently gained momentum in the research community. Due to the open nature of ad hoc networks and their inherent lack of infrastructure, security exposures can be an impediment to basic network operation and countermeasures should be included in network functions from the early stages of their design.[14] Security solutions for MANET have to cope with a challenging environment, including scarce energy and computational resources and lack of persistent structure to rely on for building trust.

These such topology control algorithms are the way to achieve a main goal i.e. a control on the energy of mobile ad-hoc networks in other words we can say that through these algorithms and protocols we can almost achieve the goal of energy consumption in mobile ad-hoc network. As it is very clear that the nodes are always in one of three state the transmitting state when the two nodes are able to share an information, the sleep state where each and every node in the network do not share any information or we can say that a passive state and the third state is receiving state where on node transmit packet and the other is receiving. All three states are power consuming in a basic ratio, but more energy can be saved in an ideal state of node where most of the nodes are in a sleep state in a network.[13] So by using such algorithms and protocols we are able to conserve energy or power in mobile ad-hoc networks. This gives the long life-time to the nodes in the network. So the choosing of protocol and algorithm is depends upon particular kind of network. Sometimes mobility of the node of the network is high and sometimes is low but the energy of the node is our prime concern.

6. REFERENCES

[1] Maleki, M.; Dantu, K.; Pedram, M, "Lifetime Prediction Routing in Mobile Ad hoc Networks," wireless communication and Networking, pages 1185-1190, vol 2, IEEE 2003.

[2] Senouci, S.-M., Pujolle, G., "Energy Efficient Routing in Wireless Ad hoc Networks", IEEE International Conference on Communications, pages 4057- 4061, vol.7, June 2004.

[3] Sanjay Kumar Dhurander1, Sudip Mishra, Mohammad S. Obaidat, Vikrant Bansal, Prithvi Raj Singh and Vikas Punia "EEAODR: an energy efficient ad-hoc on-demand routing protocol for mobile ad-hoc network", international journal of computer ad-hoc systems, pages:789-817, issue no.22, 2009.

[4] Suvarna P Bhatsangave and V.R. Chirchi, "OAODV routing algorithm for improving energy efficiency in mobile ad-hoc network", International journal of computer applications, pages-0975-8887, Volume 51, August 2012.

[5] R.Madanmohan, K.Selvakumar, "Adapting power routing in MANETS", Journal of computing, pages-2151-9617, Volume 8, Issue 4, no.2, July 2011.

[6] S.Mamechaoui, F.Didi and G.Pujolle, (2012) "Energy Saving for wireless mesh network", MESH 2012: IARIA, 2012, ISBN: 978-61208-210-3.

[7] S.Mamechaoui, F.Didi and G.Pujolle, (2012), Power Conservation for Wireless Mesh Network", 2012 IEEE 24th International Conference on Microelectronics (ICM).

[8] M. Krunz, A. Muqattash, and S. J. Lee, "Transmission power control in wireless ad hoc networks: challenges, solutions, and open issues," IEEE Network Magazine, vol. 18, no. 5, pp. 8–14, 2004. View at Google Scholar

[9] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: jointly optimal congestion control and power control," IEEE Journal on Selected Areas in Communications, vol. 23, no. 1, pp. 104–116, 2005. View at Publisher · View at Google Scholar · View at Scopus

[10] J. Gomez and A. T. Campbell, "Variable-range transmission power control in wireless ad hoc networks," IEEE Transactions on Mobile Computing, vol. 6, no. 1, pp. 87–99, 2007. View at Publisher · View at Google Scholar · View at Scopus

[11] Papadimitratos P, Haas Z, Samar P. The Secure Routing Protocol (SRP) for Ad Hoc Networks. Internet-Draft, draft-papadimitratos-secure-routing-protocol-00.txt, December 2002.

[12] [12] Zapata M, Asokan N. Securing Ad Hoc Routing Protocols. In Proceedings of the ACM Workshop on

[13] Wireless Security, September 2002, 1-10.

[14] Cano J, Kim D. Investigating Performance of Power-aware Routing Protocols for Mobile Ad Hoc Networks. IEEE MobiWac, October 2002.

[15] Chakrabarti S, Mishra A. QoS Issues in Ad Hoc Wireless Networks. IEEE Communication Magazine, February 2001; 142-148.

❑❑❑

PRESERVING CONFIDENTIALITY OF DATA IN CLOUD COMPUTING USING ENCRYPTION

[1]Shashi Teotia* and [2]Dr. Anuranjan Mishra
[1]Research Scholar, CSE Department, Noida, International University, G.Noida, Uttar Pradesh, INDIA.
[2]Prof. & Director, Accurate Institute of management and Technology, G.Noida, Uttar Pradesh, INDIA.
*Corresponding Author Email ID: teotia.shashi@gmail.com

ABSTRACT

In cloud computing, there is main concern is security. Every organization and researcher is focused on security because in the cloud , data is moved to the remote location. As there is lots of advantages of cloud infrastructure like ease maintenance, reduced cost and increased the profit. When the data is uploaded or stored to the cloud all use the standard method of encryption to secure the operation and storage of data of user or client. But to stored data on remote servers , cloud providers need to access the raw data. In our paper, we propose the most secure encryption method for data and present a framework.

Keywords: Cloud Computing, Security, Encryption, Access Control, Confidentiality.

1. INTRODUCTION

In computer industry, we asked the main question. How to be sure that if data centers of cloud provider attack by attacker , our data would be safe, not be stolen or re-used. And second one is that how our data will be confidential from unauthorized users.

The main method or technique is to encrypt the data before send or stored in cloud. The client will need to provide the key to server , that to decrypt the before the execution.

The primary attribute of Quality of Service is security and the cloud service provider has to give full assurance of security in terms of confidentiality, accessibility, privacy and integrity. Among the factors privacy is a primary and uncompromisable factor of security[4] . Encryption is the way to secure the data in the untrusted cloud server. Most of Encryption methods currently available had no effect on real time cloud applications. The possibility of their use in critical cloud application is limited. Thus we categorize different encryption algorithms based on their usability and adaptability .using Attribute Based Encryption (ABE) . The attribute based encryption is best way to secure when compared to other encryption types like Role based access as it has the capability to restrict access based on roles. As a result it is appropriate only to the small scale applications. The ABE is overhead in terms of data retrieval.

The term encryption refers to converting the original data into human unreadable form (encoding). The conversion of the encoded data into original form is known as decryption. By encrypting the data only the authorized person can decode the original data. Thus data confidentiality is achieved by the encryption. There are many encryption algorithms currently available and has its own advantages. The attribute based encryption is a proven algorithm for cloud computing environment[7].The limitations of some of attribute based encryption method are to be analysed. Attribute based encryption generally involves encrypting the attributes neither encrypting the whole data. Encryption in ABE is easy and secure and inexpensive compared to other encryption discussed. The ABE is secure because the encrypted data contains the attributes rather than the data. In case of any malicious attacks the data never is leaked. The limitation of the attribute based encryption is decryption of data is expensive [7].The attribute based encryption makes the application to be secure .the performance of the ABE is high compared to other encryption methods. Thus attribute based encryption is the solution to all cloud applications in future. The following evaluation required to be considered.

1. *Data confidentiality:* Any unauthorized party (including the cloud server) should not learn any information about the encrypted data files.
2. *Fine-grained access control:* For users in the same group or different groups, each user can be associated with different access rights which will make the scheme more reliable and efficient as a real life solution.
3. *Scalability:* The system should have the ability to work efficiently even when the number of authorized user increases.
4. *User accountability:* When an authorized user becomes dishonest and shares his/her attribute

private key with some unauthorized users, he/she should be held accountable.

5. *Collusion resistant:* The system's users should not be able to combine their attributes to decrypt the encrypted data files.

6. *Ciphertext size:* This refers to the size of the generated file after running the encryption algorithm on the original plaintext data file.
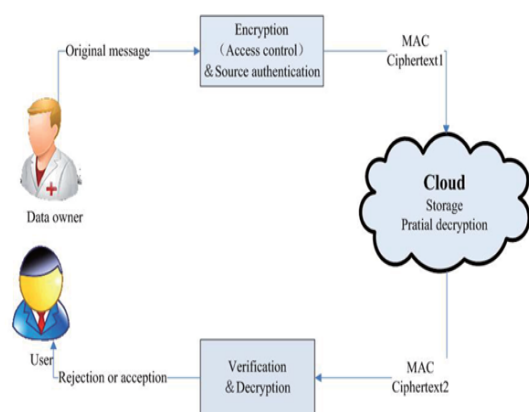

Fig. 1: Encryption

2. RELATED WORK

In Cloud Computing, the data that is stored in cloud is in unknown place. So first job is that data we stored in cloud must be secured. In order to maintain security, data location must be secret. When we talk about cloud, we state that untrusted cloud server. The above mentioned encryption techniques, The ABE are proposed to solve the complex control mechanism over encrypted database.

In 2009, Keita Emura et al, discuss with length of cipher text and consistent length of quantity match processing as a part of this plan is AND gate on multi value attribute. ABE plan which delivers constant size ciphertexts .

In 2012, AijinGe et al use secure CP-ABE scheme that support flexible threshold access structure for constant cipher text length which support threshold and AND policy.

Chase[13]proposed a multi-authority attribute based encryption framework to conquer the downsides of single attribute based framework.

ABE is public key based one to many encryption that decrypt the cipher text only if the private key associated with the user matches with public key and master secret key. The decryption of data takes place directly by the server itself. Thereby performance is increased with effective encryption.

In 2011 green et al proposed the concept of outsourcing the decryption for ABE ie user has to decrypt the data by him. The ABE with outsourced decryption overcomes the

limitation of waters and it assures security from malicious attackers. The cipher text is decrypt only with the public key matches with the user private key. The green et al algorithm modifies the water algorithm with transformation key and retrieving key. The original data is compared with partially encrypted data to achieve confidentiality of the data.

In 2013 Junzuo Lai et al proposed the ABE outsourced decryption 7. This overcomes the limitation of green et al with verifiable outsourcing of data. The proposed algorithm matches the cipher text with the decrypted cipher text. Thus verification of data is the main advantage of this algorithm. The proxy re- encryption is used for decryption of the ciphers. And size of ciphers also very small in size. The performance of the proposed system is relatively high. Thus cloud is ready for mission critical applications with outsourced decryption of data.

3. CHALLENGES

Data processing in distributed systems represents a challenge from several views. First, the scalability in environments consisting of diverse organizational entities, cross-layered rules and policies might raise issues of efficiency and interoperability. As the second, the handling of data and Privacy identifiable information should be performed on a privacy and legal conforming way, enabling users to provide informed consent and revocation.

The collaboration in multi-layer federated environments requires the storage, sharing and processing of data that exhibits various levels of sensitivity, raising the issues of conformance with security and legislative requirements or obligations. Depending on levels of trustfulness between cooperative environments, the appropriate trade-off should be applied considering the dimensions of data security, its processibility by the target system, and overall efficiency. the approaches that enable data sharing and processing in outsourced environments that support different levels of trust and access granularity both perspectives that enable the data sharing using advanced encryption schemas, as well as using novel architectures that deliver cryptographic services on various platforms and protocols.

4. FUTURE WORK

In our proposed work we will work on cloud architecture, that provide the security of data and maintain the confidentiality and provide a secure framework for working on public cloud. We will work on the following points:

Authentication: In this, we work that in our architecture only authenticated user can upload or download data. The only administrator allocated the space to user on the request basis.

Encryption and Segmentation: The data, that is received

from one cloud provider in encrypted using a secure algorithm. And encrypted data is divided into chunks. That will be part of our framework.

Decryption: The encrypted data that is received from cloud, would be de-segmented. And convert into readable form.

Downloading: In our future framework, the user request for data, and file is received from administrator is decrypted data.

5. CONCLUSION

We have classified cloud application based on the risk involved and with the parameters of suitable encryption methods. In our work, we discuss the issue of data security and processing in cloud environment. Focus of our work is with a scenario that considers various cloud services and different infrastructure of public use. Our main concern is that we can maintain the confidentiality of data in public cloud .In our future work, our aim is to integrate the public cloud services and the cryptographic approach for securing the data. Our aim to provide the transparent and security policy as well as cryptographic scheme for secure data sharing and maintain confidentiality.

6. REFERENCES

[1] A. Shamir, How to share a secret, Communications of the ACM 22 (1979) 612–613.

[2] Bojan Suzic, Andreas Reiter Florain Reimair et al "Secure data sharing and processing in hetrozenous clouds"., Elsevier , pp. 116-126, 2015

[3] R.J. Hwang, C.C. Chang, An on-line secret sharing scheme for multi-secrets, Computer Communications 21 (1998) 1170–1176.

[4] Arshad, J, Townend, P, Jie Xu.Quantification of Security for Compute Intensive Workloads in Clouds Parallel and Distributed Systems (ICPADS), 2009 15th International Conference on , vol., no., pp.479,486, 8-11 Dec. 2009.

[5] Saizhi Ye, Guoxiang Yao, Quanlong Guan, A multiple secrets sharing scheme with general access structure, in: 2009 International Symposium on Intelligent Ubiquitous Computing and Education, pp. 461–464.

[6] Saravan kumar N, Rajya Lkashmi G.V Balamurugan B "Enhanced Attribute Based Encryption for Cloud Computing", Elsevier , pp. 689-696, 2015

[7] Junzuo Lai,Deng, R.H, Chaowen Guan, Jian Weng,.Attribute-Based Encryption With Verifiable Outsourced Decryption Information Forensics and Security, IEEE Transactions on , vol.8, no.8, pp.1343,1354, Aug. 2013.

[8] Nikita Gorasia, RR Srikanth, Nishant Doshi, "Improving security in multi authority Attribute Based Encryption with Fast Decryption" Elsevier, pp. 632-639,2016

[9] Zhen Liu, Xiaoyuan Yan, Ping Wei, Haiyan Xiao, Public verifiable algorithm of threshold secret sharing with short share, Journal of Computer Applications 9 (2009) 2363–2365.

[10] N. Bourbaki, Elements of mathematics, in: Algebra: Modules. Rings. Forms, vol. 2, Addison-Wesley, 1975 (Chapter 4,5,6)

[11] Mihai Christodorescu, Reiner Sailer, Douglas Lee Schales, Daniele Sgandurra, and Diego Zamboni. 2009. Cloud security is not (just) virtualization security: a short paper. In Proceedings of the 2009 ACM workshop on Cloud computing security (CCSW '09). ACM, New York, NY, USA, 97-102.

[12] Dropbox news. https://www.dropbox.com/news.

[13] Rafail Ostrovsky, Amit Sahai, and Brent Waters. 2007. Attribute-based encryption with non-monotonic access structures. In Proceedings of the 14th ACM conference on Computer and communications security (CCS '07). ACM, New York, NY, USA, 195-203.

[14] S. Chatterjee and A.Menezes. On cryptographic protocols employing asymmetric pairings—The role of revisited, Discrete Appl. Math., vol. 159, no. 13, pp. 1311–1322, 2011.

❑❑❑

NEW DYNAMIC METRICS SUITE TO MEASURE COMPLEXITY OF COMPONENT BASED SOFTWARE

Anjali Chaudhary

Assistant Professor, CSE Department, Noida International University, G.Noida, Uttar Pradesh, INDIA.
Email Id: anjali.chaudhary22@gmail.com

ABSTRACT

This paper presents dynamic metrics of CBS at three levels, these levels are method level, component level and system level. Parameters used in proposed metrics are active components, interfaces used to interact with other components, execution time of functionality and methods in components. Further, these metrics are validated against Weyuker's set of properties. The result of validation provides a valid means to measure issues regarding the execution of components.

Keywords : CBS, Metrics, EMD, Dynamics.

1. INTRODUCTION

Component Based Software(CBS) is an effective branch of software engineering, which came into existence due to the failure of insufficient use of reusability in object-oriented programming. CBS works on the concept of reusability. Software reuse [1] is the process of creating software systems from existing software rather than building them from scratch. Reusability is helpful in reducing development time and cost of any software. Reusability is achieved in CBS by a software that is known as "component". A component communicates with other components through interfaces. It basically have two types of interfaces one is provide an interface by which component provides services to other components and second is require interface by which a component takes services from other components. According to Szyperski [2, 3] "A software component is a unit of composition with contractually specified interfaces and explicit context dependencies only. A software component can be deployed independently and is subject to composition by third party"

For a good quality software, it is very important to measure the quality of the process and the product itself, for this purpose there should be some measure, so in this field of software engineering metrics plays a very important role. Various [4, 5] attributes of software such as complexity, reusability, maintainability, testability, etc. can be judge and predict with the help of metrics. Among these attributes complexity is a major attribute which affects all other attributes of the software. High complexity leads to more efforts in testing, maintenance and reusability. There are many inherent differences [6] in CBSS and non-CBSS so the traditional software metrics are inappropriate for CBSS. It becomes necessary to quantify the various characteristics of components: before integration and after integration. To measure the complexity of components there are two aspects of component which are static and dynamic measurement of code. Static metrics measures complexity of a component before integration using component packing density and component interaction density metrics. Dynamic metrics are used to measure the complexity of a component after integration at the time of execution in a CBS. Dynamic metrics are useful to determine super component and to evaluate the degree of utilization of various components [2].

2. PROPOSED CBS DYNAMIC METRICS

Proposed metrics measure the complexity of CBS at three levels, these levels are method level, component level and system level.
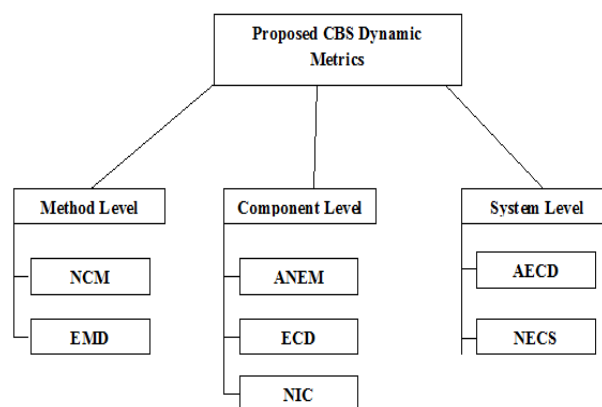

Fig. 1: Proposed CBS Dynamic Metrics

Method level metrics includes Number of Cycle in

Method (NCM), Executed Method Density (EMD). Component level metrics includes Average Number of Executed Methods (ANEM), Executed component Density (ECD) and Number of Interaction of Component (NIC). System level includes Average Executed component Density (AECD), Number of Executed Component in the System (NECS).

The parameters used in proposed metrics are active components, interfaces used to interact with other components, execution time of functionality and methods in components. So firstly complexity of a CBS is measured at method level considering executing methods in a CBS, further complexity is measured at component level using executed components in a CBS, and finally system level complexity is measured.So the formulas for these metrics are as follows:

Method Level
Number of Cycle in Method (NCM)
   NCM
$=\sum_{i=1}^{n}$ Number of Cycle of Method within a Loop(i). This metric is about how many times a method executes within a loop. It is about summation of all methods cycle executing within a loop. This metric is very useful as with the help of NCM complexity of an individual component can also be measured.

Executed Method Density (EMD)
EMD is the ratio of number of methods with internal call to the total number of methods.

$$EMD = \frac{Number\ of\ Methods\ with\ Internal\ Call}{Total\ Number\ of\ Methods}$$

In this metric only those methods are considered which are calling methods of the same component.

Component Level
Average Number of Executed Methods (ANEM)
ANEM is the ratio of number of methods (external calling) executed to total number of methods.

$$ANEM = \frac{Number\ of\ Methods\ Executed}{Total\ Number\ of\ Methods}$$

Here only those methods are considered which are calling those methods which are defined in other component. High value of this metric indicates high coupling as coupling factor is about how much one component is dependent to other component.

Executed component Density (ECD)
ECD is the ratio of total number of executed components to total number of available components.

$$ECD = \frac{Number\ of\ Executed\ Components}{Total\ Number\ of\ Components}$$

Number of executed components indicates total number of component participated to complete a particular functionality. Components interact with each other using require and provide interfaces. Thus for an application

there may be different set of ECD depending on the functionality used by the user.

Number of Interaction of Component (NIC)
NIC is the ratio of number of interactions during execution of components to total execution time.

$$NIC = \frac{Number\ of\ Interactions\ during\ Execution\ of\ Components}{Total\ Execution\ Time}$$

Here number of interactions indicates total interfaces used by components to complete a particular functionality and total execution indicates total time taken by components to complete that particular functionality. So there may be different possible set of NIC of an application.

System Level
Average Executed component Density (AECD)
AECD is the ratio of sum of ECDs to total execution time.

$$AECD = \frac{ECD1+ECD2+\cdots ECDn}{Total\ Execution\ Time}$$

Here numerator is the sum of all ECDs of different functionalities used by user in a particular application and total execution time is the sum of all execution time taken by the components to complete a particular functionality for which ECD is computed at component level.

Number of Executed Component in the System (NECS)
NECS is the total number of components which are executing for a particular functionality in an application.

NECS= Maximum Executed Component

There may be different set of NECS for a particular application as different functionalities would have different number of executed components.

3. EVALUATION OF PROPOSED METRICS
To evaluate, proposed metrics are applied on the code of a website that is Online Shopping Management system. According to the functionality of the software, component interaction diagram (Figure 2) is made in order to attain necessary parameters of formulas.
Component Interaction diagram has total twelve components. Each component can have two types of interfaces one is require interface takes input from other components and second is provide interface which gives functionality to other components.

Method Level
Number of Cycle in Method (NCM)

NCM $\sum_{i=1}^{n}$ Number of Cycle of Method within a Loop(i)
To find out the values of NCM, formula is applied on add to cart component which is one of the component of shopping management.In the code there are two methods qty_TextChanged() and GetTotal(). qty_TextChanged() is called when user change the quantity of an item of the cart

items and GetTotal() is called when user delete any item from the cart items.Each method has one foreach loop which depends on the number of items in the cart items. So there may be several possible values of NCM depending on number of items in cart item.

Case 1: If cart item has 3 items, and user change the quantity of an item and deletes one item from the cart item. Then qty_TextChanged() will be called 3 times and GetTotal() will be called 2 times . So the value of NCM will be:

$$NCM= 3+2$$
$$NCM=5$$

Case 2: If cart item has 3 items and user only change the value of an item only.Then qty_TextChanged() will run 3 times and GetTotal() will not run.

So the values of NCM will be:

$$NCM= 3+0$$
$$NCM=3$$

Fig. 2: Component Interaction Diagram

Executed Method Density (EMD)

$$EMD = \frac{\text{Number of Methods with Internal Call}}{\text{Total Number of Methods}}$$

To calculate the value of EMD the formula is applied on user login and product gallery components

For User Login component:
Total methods-3
Methods with internal call-3

For Product gallery component:
Total methods-9
Methods with internal call-5

So the value of EMD will be:
$$EMD= \frac{3+5}{3+9}$$
$$EMD= 0.66$$

Component Level
Average Number of Executed Methods (ANEM)

$$ANEM = \frac{\text{Number of Methods Executed}}{\text{Total Number of Methods}}$$

To calculate the value of ANEM again the formula is applied on user login and product gallery components.

For User Login component:
Total methods-3
Methods with external call-0

For Product gallery component:
Total methods-9
Methods with external call-4
So the value of ANEM will be: $ANEM= \frac{0+4}{3+9} = 0.333$

Executed Component Density

$$ECD = \frac{\text{Number of Executed Components}}{\text{Total Number of Components}}$$

To calculate the value of ECD different possible cases are taken which can be used by user.

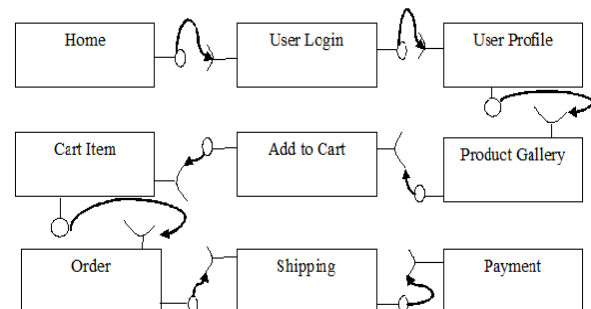**Case 1:** If user login successfully and purchase some items.

Fig. 3: Component Interaction Diagram if user login and purchase some item

Total Components= 12 (*Refer figure 2*)
Executed Component=3 (*Refer figure 5*)

$$ECD_3 = \frac{3}{12}$$

$$ECD_3 = 0.25$$

Number of Interaction of Component (NIC)

$$NIC = \frac{\text{Number of Interactions during Execution of Components}}{\text{Total Execution Time}}$$

To calculate the value of NIC different possible cases are taken which can be used by user.

Case1: If user performs successful login and go to user profile. In this case three components interact with each other and two interactions are used.

Table 1: Execution Time

EventClass	TextData	ApplicationName	NTUserN...	LoginName	CPU	Reads	Writes	Duration	ClientProcessID	SPID	StartTime	EndTime
ExistingConnection	-- network protocol: LPC set quoted...	.Net Sql...	anjali	anjal...					4596	53	2014-04-24 16:14:03...	
Audit Logout		.Net Sql...	anjali	anjal...	0	500	2	368213	4596	52	2014-04-24 16:13:58...	2014-04-24 16:20:06..
RPC:Completed	exec sp_reset_connection	.Net Sql...	anjali	anjal...	0	0	0	0	4596	52	2014-04-24 16:20:06...	2014-04-24 16:20:06..
Audit Login	-- network protocol: LPC set quoted...	.Net Sql...	anjali	anjal...					4596	52	2014-04-24 16:20:06...	
SQL:BatchStarting	select count(*) from tbl_login where...	.Net Sql...	anjali	anjal...					4596	52	2014-04-24 16:20:06...	
SQL:BatchCompleted	select count(*) from tbl_login where...	.Net Sql...	anjali	anjal...	0	68	0	37	4596	52	2014-04-24 16:20:06...	2014-04-24 16:20:06...
SQL:BatchStarting	select uid from tbl_login where unam...	.Net Sql...	anjali	anjal...					4596	52	2014-04-24 16:20:06...	
SQL:BatchCompleted	select uid from tbl_login where unam...	.Net Sql...	anjali	anjal...	0	6	0	1	4596	52	2014-04-24 16:20:06...	2014-04-24 16:20:06...
SQL:BatchStarting	select fname,lname,gender,email,city...	.Net Sql...	anjali	anjal...					4596	51	2014-04-24 16:20:06...	
SQL:BatchCompleted	select fname,lname,gender,email,city...	.Net Sql...	anjali	anjal...	0	58	0	10	4596	51	2014-04-24 16:20:06...	2014-04-24 16:20:06...
Trace Stop											2014-04-24 16:20:10...	

Table 2: Execution Time

EventClass	TextData	ApplicationName	NTUser...	LoginName	CPU	Reads	Writes	Durati...	ClientPr...	SPID	StartTime	EndTime
Trace Start											2014-04-24 16:13:47...	
Audit Login	-- network protocol: LPC set quote...	.Net SqlCli...	anjali	anjali...					4596	52	2014-04-24 16:13:58...	
SQL:BatchStarting	select pid,productname,cdate from t...	.Net SqlCli...	anjali	anjali...					4596	52	2014-04-24 16:13:58...	
SQL:BatchCompleted	select pid,productname,cdate from t...	.Net SqlCli...	anjali	anjali...	0	261	0	51	4596	52	2014-04-24 16:13:58...	2014-04-24 16:13:58..
Audit Login	-- network protocol: LPC set quote...	.Net SqlCli...	anjali	anjali...					4596	53	2014-04-24 16:14:03...	
SQL:BatchStarting	insert into tbl_products(productnam...	.Net SqlCli...	anjali	anjali...					4596	53	2014-04-24 16:14:03...	
SQL:BatchCompleted	insert into tbl_products(productnam...	.Net SqlCli...	anjali	anjali...	0	24	1	3	4596	53	2014-04-24 16:14:03...	2014-04-24 16:14:03..
SQL:BatchStarting	select pid,productname,cdate from t...	.Net SqlCli...	anjali	anjali...					4596	53	2014-04-24 16:14:03...	
SQL:BatchCompleted	select pid,productname,cdate from t...	.Net SqlCli...	anjali	anjali...	0	2	0	0	4596	53	2014-04-24 16:14:03...	2014-04-24 16:14:03..
Trace Stop											2014-04-24 16:14:10...	

Number of Interactions=2 (*Refer figure 4*)
Execution Time= 37+1+10=48 ns (*Refer table 1*)

$$NIC= \frac{2}{48}$$

NIC= 0.041

Case 2: If admin logins and update product gallery. In this case also three components interact.

Number of Interactions=2 (*Refer figure 5*)
Execution Time= 51+3=54 ns (*Refer table 2*)

$$NIC= \frac{2}{54}$$

NIC = 0.037

System Level
Average Executed component Density (AECD)

$$AECD = \frac{ECD1+ECD2+\cdots ECDn}{\text{Total Execution Time}}$$

ECD_1=0.75, ECD_2=0.25, ECD_3=0.25

The values of ECD_1, ECD_2 and ECD_3 are calculated at method level.
Execution time for the ECD_1= 16
Execution time for the ECD_2= 48
Execution time for the ECD_3= 54

$$AECD = \frac{0.75+0.25+0.25}{16+48+54} = 0.010$$

Number of Executed Component in the System (NECS)
NECS= Maximum Executed Component in the system

There may be different set of NECS for a particular application as different functionalities would have different number of executed components.

Case 1: If user interacts with the system and purchase some items

Active Components=9 (*Refer figure3)*
NECS= 9

Case 2: If admin interacts with the system

Active Components=3 (*Refer figure 5*)

NECS=3

4. VALIDATION OF PROPOSED METRICS

Weyuker has proposed an axiomatic framework for evaluating complexity measures. Proposed metrics are measured on these properties.

Property 1: There are programs/components P and Q for which M (P) ≠M (Q).

Method Level: NCM satisfies this property as different components can have different methods with different number of executing cycles. EMD also satisfies this property as different components can have different number of methods with internal call.

Component Level: ANEM satisfies this property as different components can have different number of methods with external call.ECD also satisfies as two different programs of different functionality can have distinct number of executed component and thus both have differ vales of metrics. Like these metrics NIC also satisfies as two components can have different execution time and interfaces.

System Level: AECD and NECS are satisfying the property as two systems can have different values of active components and execution time.

Property 2: If C is non-negative number, then there are finitely many programs/components P for which M (P) = c.

Method Level: NCM satisfy the property as two components can have same value of execution cycle of methods. Similarly EMD satisfies the property

Component Level: ANEM also satisfying the property as EMD and NCM. If two applications have similar number of total components, and executed components then the value of ECD would be same so this satisfy the property. Similarly NIC also satisfying this property.

System Level: NACS satisfying the property as it is possible to have two application with similar number of active components. It is not always possible to have same values of AECD of two applications.

Property 3: There are distinct components/programs P and Q for which M (P) = M (Q).

Method level: NCM satisfying the property as two components can have similar number of execution cycles. Similarly EMD also satisfies the property as two components can have similar number of methods with internal call.

Component Level: ANEM is satisfying the property as two components can have same number of methods with

external calling finally have same values of metrics. Similarly ECD and NIC also satisfy the property.

System Level: AECD and NACS satisfy the property as two different systems can have similar number of active components and same execution time.

Property 4: There are functionally equivalent components/programs P and Q for which M (P) ≠M(Q).

Method Level: NCM satisfy the property as two components can be developed in different programming languages then can have different number of execution cycles of methods Similarly EMD satisfies the property.

Component Level: ANEM, ECD and NIC satisfy the property as components are language dependent and each language has different complexity.

System Level: AECD also satisfy the property for the same reason of language independency. But NECS does not satisfy this property if two components have same functionality then they would have similar numbers of active components.

Property 5: For any program/component bodies P and Q, we have M(P)<=M(P;Q) and M(Q)<=M(P;Q).

Method Level: It is possible to have more number of cycles if two components are assembled so NCM satisfies this property. Similarly EMD also satisfies the property.

Component Level: ANEM does not satisfy the property as number of external calls will decrease. ECD satisfy the property as the value of total executed components will increase. NIC also satisfies the property.

System Level: AECD and NECS satisfy this property as assembled component can have more active components and execution time.

Property 6: There exist program/component bodies P, Q and R such that M(P)=M(Q) and M(P;R) ≠ M(Q:R)

Method Level: NCM satisfies the property as two components of equal values of NCM may have different values of NCM after assembling it with other component.

EMD also satisfy the property they can have different number of internal calling.

Component Level: ANEM satisfies the property as it may decrease external calling. ECD and NIC also satisfy the property as both combination can have different number of interfaces and execution time and active components.

System Level: AECD and NACS also satisfying the property as both the combinations may differ in number of active components and in execution time.

Property 7: There are program/component bodies P and

Q such Q is formed by permuting the order of statements of P and M (P) ≠ M (Q).

Permutation of the order of statement does not affect the metric value so no metric satisfy this property.

Property 8: If P is renaming of Q, then M(P)=M(Q).

Renaming of component does not affect proposed metrics so all metrics satisfy this property. Proposed metrics are only based on methods, interfaces, execution time and number of active components.

Property 9: There exist program/component bodies P and Q such that M (P) + M (Q)<M(P;Q).

Method Level: If two different components are assembled with certain number of cycles of methods then it is not always possible that assembled component will have increased value of NCM. So NCM does not satisfy this property. Like this EMD does not satisfy this property.

Component Level: If two components are assembled with certain number of methods then it is not necessary that resulting component would have increased complexity.

So ANEM does not satisfying this property. ECD satisfy this property because assembly of two components will decrease in number of total components and hence complexity will increase.NIC satisfying the property as total execution time will decrease.

System Level: AECD satisfies as execution time will decrease and NACS satisfies the property as combination of components will result, increment in the number of active components.

Table 3

Metrics	Weyuker's Property								
	1	2	3	4	5	6	7	8	9
NCM	✓	✓	✓	✓	✓	✓	x	✓	x
EMD	✓	✓	✓	✓	✓	✓	x	✓	x
ANEM	✓	✓	✓	✓	x	✓	x	✓	x
ECD	✓	✓	✓	✓	✓	✓	x	✓	✓
NIC	✓	✓	✓	✓	✓	✓	x	✓	✓
AECD	✓	x	✓	✓	✓	✓	x	✓	✓
NECS	✓	✓	✓	x	✓	✓	x	✓	✓

5.   CONCLUSION

There are so many complexity metrics proposed by different researchers for CBSS. But there are very few metrics which measures the complexity of a component during execution, when it interacts with other components. It becomes necessary to measure the complexity of a component during execution to estimate the dynamic behavior of the whole application. So this study leads the development of a dynamic metric to measure the complexity of a component during execution.

6.   REFERENCES

[1] Satish Kumar Soora (2014) "A Framework for Software Reuse and Research Challenges." International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 10.

[2] Lakshmi Narasimhan V, Bayu Hendradjaya (2005) "Theoretical Considerations for Software Component Metrics" In Proc World Acad Sci, Eng Technol. 2005. ISSN 1307-6884.

[3] Clemens Szyperski (1998), "Component Software; Beyond Object-Oriented Programming" Addisonsley.

[4] Usha Kumari and Shuchita Upadhyaya(2011), "An Interface Complexity Measure for Component-based Software Systems," International Journal of Computer Applications", Vol 36, No.1.

[5] P. K. Suri and Neeraj Garg (2009), "Software Reuse Metrics: Measuring Component Independence and its applicability in Software Reuse", International Journal of Computer Science and Network Security, VOL.9, No.5.

[6] Pooja rana, Rajendra singh "International Journal of Emerging Research in Management &Technology" ISSN: 2278-9359 (Volume-3, Issue-11).

❑❑❑

MAINTENANCE APPROACH TO REDUCE EFFORTS IN SOFTWARE DEVELOPMENT LIFE CYCLE

[1]Mahtab Alam and [2]Kunwar Babar Ali*
[1,2] Assistant Professor, Department of Computer Science,
Noida International University, G.Noida, Uttar Pradesh, INDIA.
*Corresponding Author Email ID: babar.ali@niu.edu.in

ABSTRACT

The software development process is combination of several phases and maintenance is one of the major components of it. It has been considered that about 70% or more of the total software development budget cost is expending in maintenance of a software application. Software maintenance is a process of activities when an application after deployment to modify an existent software application due to some error or in anticipation of future problems. A number of works have been done to reduce the maintenance cost and effort of an application; herewith we are proposing an integral approach to reduce the maintenance cost while designing the software application. In this particular research work we are proposing different approaches at design phase to reduce the cost and effort of all kinds of software maintenance during the entire life cycle of an application.

Keyword: Software Maintenance, Scalability, Reliability, Mean Time between Failure (MTTBF), Mean Time To Recover (MTTR).

1. INTRODUCTION

Software maintenance is a set of activity performed when software undergoes modification to code and associate documentation due to a problem or the need for improvement [1]. By the laws of software evolution, maintenance decisions are aided by understanding what improves to system overtime. We are interested in change in size, complexity, resources and ease of maintenance. Software maintenance is become a major activity in the industry. A surveys and estimate made between 1988 and 1990 suggested that an average as much as 75% of a project software budget is devoted to maintenance activity over the life of the software [2]. Software maintenance costs are the greatest cost incurred in developing and using a software system. Maintenance cost varies widely from application to application, but an average they seem to be between 2.0 to 4.0 times developments costs for large software system [3].

Software maintenance is the degree to which it can be understood, corrected, adapted and/or enhanced. Software maintenance accounts for more effort than any other software engineering activity. When the changes in the software requirement are requested during software maintenance, the impact cost may be greater than 10 times the impact cost derived from a change required during the software design. i.e. the cost to maintain one line of source code may be more than 10 times the cost of the initial development of that line[4] Maintenance in the wildest sense of post development software support, is likely to continue to represent a very large fraction of the total

system cost [3]. As more programs are developed the amount of effort and resources expanded on software maintenance is growing. Maintainability of software thus continues to remain a critical area in the software development era. Verification and Validation (V & V) for software maintenance is different from planning V&V for development efforts [5]

Maintenance may be defined by defining four activities that are undertaken after a program is released for use. First activity is the corrective maintenance that corrects uncovered an error after software is in use, Adaptive maintenance; the second activity is applied when changes in the external environment precipitate modification to software. The third activity incorporates enhancement that are requested by customers and is defined by perfective maintenance where most of the maintenance cost and efforts are spent. The fourth and last activity is preventive maintenance which is in anticipation of any future problem. The maintenance effort distributions are as follows [6]:

Activity	% Efforts
Enhancement	51.3
Adaptive	23.6
Corrective	21.7
Others	3.4

Maintainability has been defined as effort of personnel hours, errors caused by maintenance actions, scope of effort of the maintenance action and program

comprehensibility is subject to the programmer experience and performance [7].

Cost factor is an important element for the success of a project. Cost in a project is due to the requirement of hardware, software and human resources. Cost estimates can be based on subjective opinion of some person or determined through the use of models [8]. Reliability-Constrained Cost Minimization cost subject to a system reliability goal. Reliability of a system is presented as a function of component failure intensities, as well as operation profile and component utilization parameters. Let n denote the number of software components.ρ denotes the system reliability target and $\tau > 0$ be the mission time, the probability of failure free execution with respect to time interval $[0,\tau]$ to be at least ρ. We assume that $0 < \rho < 1$. the total cost (TC) of achieving failure intensities $\lambda1, \lambda2,\ldots\ldots\lambda n$ and $R(\lambda1,\lambda2,\ldots\lambda n,\tau) \geq \rho$ [9]

$\lambda i \geq 0$ for $I = 1,2,\ldots.n$

The purpose of the software cost model is to produce the total development effort required to produce a given piece of software in terms of the number engineers and length of time it will take to develop the software. The general formula used to arrive at the nominal development effort is [10]

$$PM_{initial} = c.KLOC^k$$

Where   PM = person per month
        KLOC = Thousand of line of code
        C and k are constant given the model.

Software metrics are numerical data related to software development. Metric strongly supports software project management activities. They relate to the four function of management which are as follows [11]:

- *Planning:* Metric save as a basis of cost estimating, training, planning, and resource planning, scheduling and budgeting.
- *Organizing:* Size and schedule metrics influence a project organization.
- *Improving:* Metrics are used, as a tool for process improvement efforts should be concentrated and measure the efforts of process improvement efforts.
- *Controlling:* Metrics are used to status and track software development activities for compliance to plan.

The first step on the maintainability analysis using metrics is to identify the collection of metrics that reflects the characteristics of the viewpoint with respect to which the system is being analyzed and discard metrics that provide redundant information [12].

Object oriented technologies greatly influence software development and maintenance through faster development, cost saving and quality improvement and thus has become a major trend for methods of modern software development and system modeling [13]. Class, object, method, message, instance variable, and inheritance are the basic concept of the object oriented technology [14]. Object oriented metrics are mainly measures of how these constructs are used in designed process. Classes and methods are the basic constructs for object oriented technology. The amount of function provided by object oriented software can be estimated based on the number of identified classes and metrics or its variables.

Improving the quantity and reducing the cost of products are fundamental objective of any engineering discipline. In the context of software as the productivity and quality are largely determined by the process to satisfy the engineering objectives of quality improvement and cost reduction, the software must be improved. Cost factor is the crucial aspects of project planning and managing. Cost overrun can cause customers to cancel the project and cost underestimate can force a project team to invest much of its time without financial compensation.

2. MAINTENANCE : A DIFFERENT OPINION

The maintenance software is affected by many factors, such as the availability of skilled staff, the use of standardized programming languages and inadvertent carelessness in design. Implementation and testing has an obvious negative impact on the ability to maintain the resultant software. Additionally, some software organization may become maintenance bound, usable to undertake the implementation of new projects, because all their resources are dedicated to the maintenance of old software. The opinion of Programmers, Managers and Customers are as follows:

- *Programmer's opinion:* According to programmers' opinion, a program with a high level of maintainability should consist of modules with loose coupling and high cohesiveness, simple, traceable, well structured, well documented, concurrent sufficiently commented code, well defined terminology of their variables. Furthermore, the implemented routines should be of a reasonable size, preferably less than 80 lines of code with limited fan-in and fan-out. Finally the declaration and the implementation part of each routine must be strictly separated.
- *Program Managers Opinion:* Program Manager always aims at the limitation of effort spent during the maintenance process. They also focus on the high reusability of one program.
- *Customers Opinion:* Nowadays, because of the high demand of the successful software systems and external changes, a high level of modification can be attributed to changes in requirement.

3. DESIGN CONSIDERATION: A BETTER WAY TO REDUCE COST AND EFFORTS

Several elements affect and shape the design of the application. Some of these elements might be non-negotiable and finite resources, such as time, money and workforce. Other elements such as available technologies, knowledge and skills are dynamic and vary throughout the development life cycle [15]. Analyze the high level design of a software system for the purpose of prediction with respect to change difficulty from the point of view of the testers and maintainers [16]. The decision for scalability is set in the context of a software engineering environment [17]. Although these elements influence the design of an application to some extent, the business problems dictates the capabilities application must have for a satisfactory solution, such are as follows:

3.1 DESIGN FOR SCALABILITY

Scalability is the capability to increase resources to produce an increase in the service capacity. A scalable application requires a balance between the software and hardware used to implement the application. The two most common approaches to scalability are:

3.1.1 SCALING UP

Refers to achieving scalability by improving the existing servers processing hardware. Scaling up includes adapting more memory, more or faster processes or migrating the application to a powerful computer. Typically, an application can be scale up without changing the source code. In addition the administrative efforts do not change drastically. However, the benefit of scaling up tapers off eventually until the actual maximum processing capabilities of the machine is reached as shown in figure 1.

Fig.1. Scaling Up

3.1.2 SCALING OUT

Refers to distributing the process load across more than one server. This is achieved by using multiple computers; the collection of computers continues to act as the original device configuration from the end user perspective. The application should be able to execute without needing

information about the server on which it is executing. This concept is called location transparency. It increases the fault tolerance of the application as shown in figure 2.

Table 1: Measurement Types for Calculating Availability

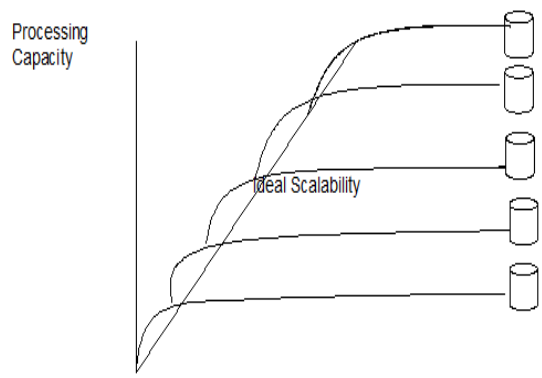| Name	Calculation	Definition
Mean Time Between Failure (MTBF)	Hours/Failure Count	Average length of time the application runs failing
Mean Time To Recovery (MTTR)	Repair Hours/Failure Count	Average length of time needed to repair and service after a failure.

Fig. 2: Scaling Out

The design has more impact on the scalability of an application than the other three factors. As we move up the pyramid, the impact of various factors decreases as shown figure 3:
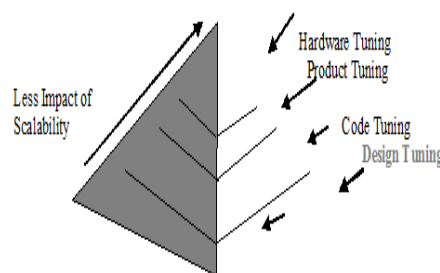

Fig. 3: Design Pyramid

To design for scalability, the following guidelines should be considered:

- Design process such that they do not waist.
- Design process so that processes do not complete for resources.
- Design processes for compatibility.
- Partition resources and activities.
- Design component for interchangeability.

3.2 DESIGN FOR AVAILABILITY

Availability is a measure of how often the application is available to handle service requests as compared to the planned run time. Availability also takes into account

repair time because an application that is being repaired is not available for use. The measurement types of availability are shown table 1.

The formula for calculating availability is:

Availability = (MTBF / (MTBF + MTTR)) X 100

The MTBF for a system that has periodic maintenance at a regular interval can be Described by [18] as follows:

$$MTBF = \int_0^{} R_T(t)dt/(1 - R_T(t)$$

Where RT (t) = exact, reliable function assuming periodic maintenance every T (hours).

Designing for availability include anticipating, detecting and resolving hardware and software failures before they result in service, errors faults, or data corruption thereby minimizing downtime.

To design for availability of an application the following guidelines should be considered:
- *Reduce Planned downtime:* Use rolling upgrades, e.g. to update a component on a clustered server, we can move the server's resources group to another server, take the server offline, update the component and then bring the server online. Meanwhile, application experiences no downtime.
- *Use Redundant array of independent disks (RAID) :* Raid uses multiple hard disks to store data in multiple places. If a disk fails, the application in transferred to a mirrored data image and the application continues running. The failed disk can be replaced without stopping the application.

3.3   DESIGN FOR RELIABILITY

The Reliability of an application refers to the ability of the application to provide accurate results. Although software standard and software engineering processes guide the development of reliable or safe software, mathematically sound conclusions that quantify reliability from conformity to standard are hard to drive [19]. Reliability measures how long the application can execute and produce expected results without failing.

The following tasks can help to create reliable application.
- Using a good architectural infrastructure.
- Including Management Information in the application.
- Implementing error handling.
- Using redundancy.

3.4   DESIGN FOR PERFORMANCE

Performance is defined by metrics such as transaction throughput and resource utilization. An application performance can be defined in terms of its response time.

To define a good performance the following steps should be taken.

- Identify project constraints.
- Determine services that the application will perform.
- Specify the load on the application.

3.5   DESIGN FOR INTEROPERABILITY

Interoperability refers to the ability to operate an application independent of programming language, platform and device. The application needs to design for interoperability because it reduces operational cost and complexity, uses existing investment and enables optimal deployment.To design application interoperability the following tasks should be considered:
- Network interoperability.
- Data interoperability.
- Application interoperability.
- Management interoperability.

3.6   DESIGN FOR GLOBALIZATION

Globalization is the process of designing and developing an application that can operate in multiple cultures and locales. Globalization involves:
- Identifying the cultures and locales that rus be supported.
- Designing features that support those cultures and locales.
- Writing code that executes property in all the supported cultures and locales.

Globalization enables to create an application that can accept, display and output information in different languages scripts that are appropriate for various geographical areas.

To design for globalization the following information should be kept in mind:
- Character classification.
- Date and Time formatting.
- Number, currency, weight and measure convention.

3.7   DESIGN FOR CLARITY AND SIMPLICITY

Clarity and simplicity are enhanced by modularity and module independence, by structured code and by top-down design and implemented among other techniques. The allocation of functional requirements to elements of code represents an important step in the design process that critically impact modifiability [20].The following guidelines for the definition of modules will have an extremely positive impact on maintainability:
- Use hierarchical module control structure whenever possible.
- Each module should do its own housekeeping as first act.

- The module should have only one entrance and exit.
- Limit module size. Up to 200 statements.
- Reduce communication complexity by passing parameters directly between modules.
- Use 'go-to-less' or structured programming logic.

3.8 DESIGN FOR READABILITY

Maintenance will ultimately result in changing the source code, understanding of thousand lines source code is almost impossible if source code is not well supported by meaningful comments. So, readability of the source code can be estimated by finding the percentage of comments lines in total code. A new factor common ration (CR) is defined as[21}

CR = LOC/LOM
LOC = Total Lines of Code
LOM = Total Lines of Commented in the Source Code.

4. RESULT AND DISCUSSION

The above practices of designing software will reduce the cost and efforts of maintaining software during the maintenance phase of the software development life cycle. Design of scalability will help to scale up and scale in of software at any time, while the design of reliability will help to develop a software which will provide the desired output. **Alam et. al** presented a software secure requirement metrics using a checklist in which all the security parameters proposed by Jan Jurjen are considered. With the help of these available checklists Degree of Secure Requirement (DSR) metrics can calculated the security concern of any proposed requirement [22][23].

5. CONCLUSION

Software Maintenance is a set of activities when software undergoes for improvement due to an existing error or in anticipation of future problems. The life span of software entirely depends on a good design. Software design provides a blue print in which the entire software built. A good design software will be less problematic, so software industries require a good design paradigm and approaches for reducing the cost and efforts of software maintenance. In this paper, we have proposed a number of design approaches for the same. In future we will develop some design tool for better ways of designing a software application.

6. REFERENCES

[1] Pressman, R., "Software Engineering. A Practical Approach", *4th Edition, Mc Graw Hill*, 1997

[2] Khan, R. A., Mustafa, K. "Assessing Software Maintenance – A Metric Based Approach", *DeveloperIQ,* March 2004.

[3] Shari Lawrence Pfleeger, "Software Engineering Theory and Practice", *2nd Edition, Pearson Education Asia 2002.*

[4] Ian Somerville "Software Engineering" *5th Edition, Addison Wesley*, 1996.

[5] Wallace, R. Dolores, Daughtrey, Taz, " Verifying and Validating for Maintainability", *IEEE Press*, 1988, pp. 41-46.

[6] Sunday, A. David, "Software Maintainability- A New 'ility',", *IEEE Proceedings Annual Reliability and Maintainability Symposium.*", 1989, pp. 50-51

[7] Sneed, M. Harry., Kaposi, Agnes., "A Study on the Effect of Reengineering Upon Software Maintainability*", IEEE Press*, 1990, pp. 91-99.

[8] Pankaj Jalote. " An Integrated Approach to Software Engineering", *2nd Edition, Narosa Publishing House.*

[9] Halender, E. Mary., " Planning Models for Software reliability and Cost", *IEEE Transaction on Software Engineering*", Vol. 24, No. 8, June 1998, pp. 420-434

[10] Ghezzi, Carlo., Jazayeri, Mehdi., Mandrioli, Dino., "Fundamental of Software Engineering", *Prentice-Hall of India Pvt. Ltd.*, New Delhi, 2002.

[11] http://sunset.usc.edu/classes/css77b-2001/metricsguide/metrics.html

[12] Muthana, S., Knotogiannis, K., Ponnambalam, K., Stacy, B., "A Maintainability Model for Industrial Software Systems Using Design Level Metrics", *IEEE Transaction on Software Engineering*, 2002, pp. 248-256.

[13] Chu, C. Williams., Lu, Chih-Wei., Chang, Chih-Hung., Chung, Yeh-Ching., Huang, Yueh-Min., Xu, Baowen., "Software Maintainability Improvement : Integrating Standards and Models*", IEEE Proceedings of the 26th Annual International Computer Software and Application Conference*", 2002.

[14] Kan, K. Stephen, "Metrics and Models in Software Quality Engineering*", 2nd edition, Pearson Education*, 2003.

[15] "Analyzing Requirements and Defining", *Microsoft.Net Solution Architecture, Prentice-Hall of India Pvt. Ltd.*, New Delhi, 2004.

[16] Briand, C. Lionel., Morasca, Sando., Baslii, R. Victor., "Measuring and Assessing Maintainability at the End of High Level Design", *IEEE Press* 1993, pp. 88-97.

[17] Han, Jun., " Designing for Increased Software Maintainability*", IEEE Press* 1997, pp. 278-286

[18] Mitchel, J. Mondro., "Approximation of Mean Time Between Failure when a system has periodic Maintainability", *IEEE Transaction on Reliability*, Vol. 51, No. 2, June 2002, pp. 166-167.

[19] Pai, J. Ganesh, Dugan, Bechta Joanne., "Enhancing Software Reliability Estimation Using Bayesian Networks and Fault Trees", *ISSRE and Chillarege Corp*. 2001.

[20] Munson, B. John., "Software Maintainability: A Practical Concern for Life Cycle Cost", *IEEE Press* 1978, pp. 54-59.

[21] Aggarwal, Krishan., Singh, Yogesh., Chabra, Jitender Kumar., "An Integrated Measure of Software Maintainability", IEEE Proceedings Annual Reliability and Maintainability Symposium, 2002, pp. 235-241

[22] Mahtab Alam, "Software Security Requirements Checklist", *International Journal of Software Engineering*, Vol-2, 2010, pp-58-68.

[23] M.U. Bokhari, Mahtab Alam , "Security Requirement for Software Quality - A Survey of Engineering Discipline", *International Journal of ICT and Management*, Vol. I, Issue 2, October 2014, pp 125-133.

●●●

FORECAST FOR A SOCIAL PATTERN EXAMINATION IN HUGE INFORMATION IN BIG DATA

¹Madan Mohan* and ²Dr. Anuranjan Mishra
¹Research Scholar, CSE Department, Noida International University, G.Noida, Uttar Pradesh, INDIA.
²Prof. & Director, Accurate Institute of Management & Technology, G.Noida, Uttar Pradesh, INDIA.
*Corresponding Author Email ID : mmphdcse@gmail.com

ABSTRACT

The development of interpersonal organizations has prompted the critical impact of example utilization of the web. This has made ready for the immense increment of the web everywhere throughout the world including those parts that couldn't get to the web previously. It has progressed toward becoming something common for the general individuals since it has been made simple for any sensible individual to get to the web. The uncommon increment in the utilization of the web is because of the expansion or the quick utilization of the regular systems that advanced to get huge volume of data crosswise over different fields inside a brief period. This can be portrayed by the gigantic measure of information that is being created so that to influence them to be able to classify this sort of information as Big Data. Consequently, the use of enormous information precise techniques in a large portion of the fields has had incredible effect and acquired changes the task of the diverse organizations, and afterward it can be certain that interpersonal organizations are not a special case to this specific observation.

Keywords : Forecast, Prediction, Social Trend, Big Data, Dataset, Data Analysis, Machine Learning.

1. INTRODUCTION

Big Data Analytics has created a significant change in the way that businesses and people perceive data. Big data is expensive for the rival organizations since there is a particular way in which the decisions are to be undertaken [1]. An individual myth claims that any huge volume of data can be said to be big data, but this is not it since the data that is generated by systems and the one that has the ability to fulfill the constraints that are imposed in the dimensions of volume, velocity, variety, and veracity are those that can be classified as big data. This means that velocity, veracity, volume, and variety are the attributes that make up the big data [1]. When an organization takes the initiative of generating data at a higher rate and with different characteristics, and this data is used to make well-informed decisions, then this is the big data. The processing of big data can be much different from the way the current data is being or can be handled. There are those special tools and techniques that are required so that to accomplish an effective analytical decision. The data scientists prefer that they use big data sets towards the provision of informed decisions [1]. This essay is to analyze a machine learning approach for prediction for a social trend analysis in big data.

2. BIG DATA
Big data is a word used to describe large-scale structured and unresolved data very much, it is very difficult to process data using traditional database and software technology.

Volume: The quantity of produce and stored data, Many factors contribute to an increase in volume, live streaming data and Storing transaction data, live streaming data and data collections from sensors

Variety: There are data in all types of formats: text documents, from established databases, text documents, emails, video, audio, transactions etc.,

Velocity: Means how fast the data is generated and how much data needs to be treated to meet the request.

The other two aspects need to consider respect to Big Data are Variability and Complexity.

Variability: Together with speed, the data flow may not be consistent with the periodic top of something height.

Complexity: When data comes from multiple sources, you also need to consider the complexity of the data. Before actual processing, the data must be linked, matched, cleaned and converted to the desired format.
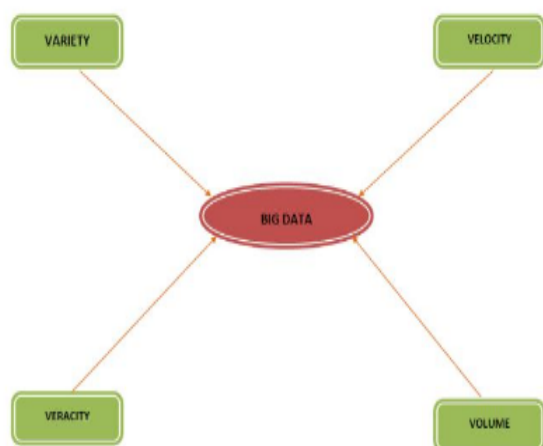
Fig. 1: The Attributes of Big Data

Hadoop: It is an open source Java-based programming framework supports the Handling large data sets in distributed computing surroundings.

It can run on thousands of terabytes of systems involving thousands of nodes. The distributed file system in Hadoop helps to achieve fast data transfer rates, and the system continues to function even in the event of some node failure. This approach reduces the risk of a total system failure, even in the case of a large number of node lack of success. Hadoop Make the calculation resolution can expand, economy, flexible and fault-tolerant. In these days so many companies are using Hadoop Framework to support applications that involve big amounts of data.

Hadoop has two main subprojects Map Reduce and Hadoop Distributed File System.

3. HADOOP DISTRIBUTED FILE SYSTEM (HDFS)

Hadoop Distributed File System is a file system that spans data across all nodes in a Hadoop cluster. It links the file systems on the local node together to make it a large file system. Hadoop Distributed File System overcomes node failures by replicating data across multiple sources, improving reliability

Big data applications
Big data applications are large-scale distributed applications, often working with big data sets. Data mining and analysis have become a problem in many areas of big data. Due to the large amount of data and large amount of data, traditional data processing applications are difficult to handle, which has led to the development of big data applications. Google's map

Big data and social network analytics
The vast generation of data has been led by the fact that there are the social networks such as Facebook, Twitter, and LinkedIn, etc. that have the paved way. For such

social media data, there is the necessity of using or applying the big data analytics so that to come up with the best or important information to the information users and the data generators [2]. The techniques and the tools used for the big data analytics include prognostic analytics, information mining, data visualization, complex SQL, synthetic intelligence as well as normal language dispensation. The investigation of both organized and formless data that is from the common networks is likely to lead to public network analytics. According to this research, it is evident that blogs and micro blogs, as well as wikis, can as well contribute to social networks analytics data sets [3]. There is where the user posted information becomes important for users, and this is called or referred as content-based analytics. Structure-based analytics happens when the relationship between entities is used for analytics. There are a million of connections that are there between the connected objects and analysis of data on the social networks, and this is expensive and intensive as well.

Two approaches must be followed, and these include the parallelization and graph databases approach [1]. A parallelization is an approach that focuses on separating a huge dataset into smaller subsets and make sure that it utilizes the computational power by the use of cloud figuring to development of the information in a parallel way. The pioneers of this approach are map reduce and Pregel which are from Google. Graph databases are well suited for networks such as social networks and transport networks where the data structure is directly used to represent the key aspect of the problem that is being analyzed [2]. The elements of the graph database have a direct pointer, and there is no need for the index lookups. Graph databases are known to be faster than relational databases especially when they are dealing with the associative data sets. They are the kind of databases that do not require expensive join operations.

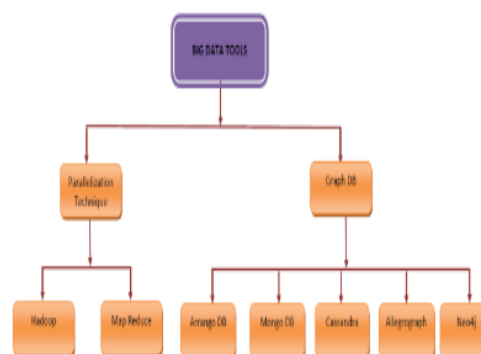**Machine learning and social network analytics**

Fig. 2: Groups of Big Data Tools

Machine learning methods are the process of teaching or showing knowledge or what to be done to any instrument such as PC, processer or the mobile devices that can be

used to learn about a particular system which has an input or the dependent variables and the desired output [3]. It is clear that any machine can perform learning and they can do this is three different modes. The three styles are overseen, unproven and the strengthening methods. The instrument learning methods can be casted off in any system to carry out and produce the necessary consequences that are required at the end, and they are fragment of extrapolative analytics and the estimating approaches [3]. The instrument learning methods are classified into verdict tree, linear as well as the regression and neutral based network.

A lot of organization are adapting to the use of social media since it has a lot of impacts, especially when making the decisions [2]. When there is the usage of public media in the making of decisions, it is more necessary to adapt to the use of massive datasets that are obtained from the social networks by the use of machine learning techniques. This makes it easy for the organizations to foresee certain situations and they will be able to make decisions appropriately according to the outcome or the results of public media analytics. The feature of the instrument techniques is the ability to iterate [2]. The iterative aspect enables the system to be independent in that it can quickly adapt to the new sets of input since the data are subjected to a variety or some datasets. The device learning techniques that have taken part in societal media examination comprise the resolution tree erudition, inexperienced Bayes, Dynamic Language Model classifier, Nearest Neighbor classifier, Supreme Entropy technique, Sustenance vector machine (SVM), linear reversion and logistic reversion, modest logistic classifier, Multilayer Perceptron, and Bayes Net [2].

Fig. 3: The Application Fields of Big Data

4.   CONCLUSION

A considerable measure of reviews has been directed on huge information that is in interpersonal organization investigation and machine learning. There is a significance in breaking down the development of online networking information that is using Big Data and the gadget instruction calculations. After all the investigation, obviously there is the need to direct further work by mix of huge information and machine learning methods since the work has augmented during recent decades.

5.   REFERENCES

[1] Boyd-Graber, K. Glasgow and J. Zajac, "Spoiler alert: Machine learning approaches to detect social media posts with revelatory information", *Proceedings of the American Society for Information Science and Technology, vol. 50, no. 1, pp. 1-9, 2013.*

[2] Pardo, "Social learning graphs: combining social network graphs and analytics to represent learning experiences", *International Journal of Social Media and Interactive Learning Environments, vol. 1, no. 1, p. 43, 2013.*

[3] Gruzd, D. Paulin and C. Haythornthwaite, "Analyzing Social Media and Learning Through Content and Social Network Analysis: A Faceted Methodological Approach", *Journal of Learning Analytics, vol. 3, no. 3, pp. 46-71, 2016.*

❑❑❑

CLOUD COMPUTING & SECURITY OF DATA

Keshav Krishna Agarwal

Professor , Computer Science & Engineering Department, SET
Noida International University, G.Noida, Uttar Pradesh, INDIA
Corresponding Author : Kka78901@gmail.com

ABSTRACT

Cloud Computing has become a necessity in present timings as most of organizations are using frequently its services. Mainly there are three concepts used in Cloud Computing i.e. IAAS, PASS & SAAS.So the main concern is its security. A rapid transition from the main servers to cloud machines is necessary for fast delivery & deployment of Cloud Computing Models. When the data are stored on cloud, there are many securities issue in this computing. In this research paper number of risks and challenges have been noticed in cloud services. Due to this reason we shall stress on how to implement security issues and conclude with the issue in cloud by secure architecture in future utilization of cloud services. Any office when want to shift their work to cloud computing there are many issues like privacy, security & fear that some personal data could be in wrong hands especially sensitive data may be available to unauthorized persons through cloud. In the cloud computing, organization have little or no visibility to storage. When we move or upload our sensitive and regulated data in cloud, issues of security must be considered.

1.   INTRODUCTION

Any office when want to move for cloud computing there are many risks like secrecy, some personal data could be in wrong hands, especially sensitive data may be available to unauthorized user through cloud. In the cloud computing, organization have little or no visibility to storage. When we move or upload our sensitive and regulated data in cloud, issues of security must be considered.

Cloud computing has become very popular in many industries &organizations. It represents a new computing which is used to computation as well as storage. The term cloud in cloud computing has come from the symbol cloud that is generally used in to show the internet in diagrams and figures. Cloud computing is empowered by the virtualization. There are lots of benefits to move cloud but when an organization moves towards cloud, the number of risks are available. In the cloud, the organization have very little or no visibility to storage and backup process of data. When sensitive data is moved into public cloud, issues of security standards and compliance should be involved like authorization, encrypted data and data loss protection.

Cloud computing is defined as cloud computing is a model for enabling, ubiquitous,convenient, on-demand network access to a shared pool of configurable computing resources network, server, storage, application and services that can be rapidly provisioned and released with minimal management efforts or service provider interaction. This cloud model is composed of five essential characteristics, three service models and four deployment model. Essential Characteristics of cloud computing includes as below.

On-demand self-service: A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

Broad network access: Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).

Resource pooling: The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter). Examples of resources include storage, processing, memory, and network bandwidth.

Rapid elasticity: Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be

appropriated in any quantity at any time.

Measured service: Cloud systems automatically control and optimize resource use by leveraging a metering capability1 at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.

The CSA has identified thirteen domains of concerns on cloud computing security.

In the cloud computing, there are three service models are available

- **Infrastructure as a Service(IaaS):** It provides the consumer with the capability to processing, storage, network and other resources. It allows the consumer to deploy and run the software which can include the operating system and applications. The user has the control over operating system, storage.

- **Software as a Service(SaaS):**It provides the capability to the consumer is to use the provider's applications running on a cloud infrastructure2. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration setting.

- **Platform as a Service(PaaS):** The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.

2. SECURITY OF CLOUD
2.1 CONFIDENTIALITY

It refers to the only authorized system have the ability to access the data. It is a term that correlated with authentication in cloud computing. Authentication is the process of establishing confidence in user identity to information system. Multitenancy refers to resource sharing in cloud. when multi tenancy presents number of privacy and confidentiality threats. Data confidentiality refers to trusting the specific application or data in a secure manner.

2.2 TRUST

Trust is the new topic in area of computer science as diverse as security and access control. The concept of trust as described "Anentity A is considered to trust another entity. when entity A believes on B, B will behave as required and expected. Trust is cloud computing is depending on deployment model. When data is deployed on public cloud, control is mitigated to the infrastructure. This introduced the number of risks and threats that is related to trust.

2.3 PRIVACYLEGAL ISSUES

It the desire of a person to control the disclosure of private data or information. Enterprises deal with the private data are required to comply with legal framework. Cloud computing has numerous legal issue or challenges towards privacy of data that is stored in multiple locations of cloud. Data stored in multiple locations also increase the risk of confidentiality and integrity. In cloud, the user's data can be on any datacenters which can be used anywhere in the world. In cloud computing, privacy is an important issue and its importance in both the terms, one is legal perspective and other one is user trust.

Whenever an individual or organization or other entity share information in the cloud, the question of privacy may arise. In cloud computing system, the relation of the user and the service provider is very complicated as compared to other kind of services. In cloud computing, cloud service provider is one type of third party that maintain information about on or behalf of other entity.
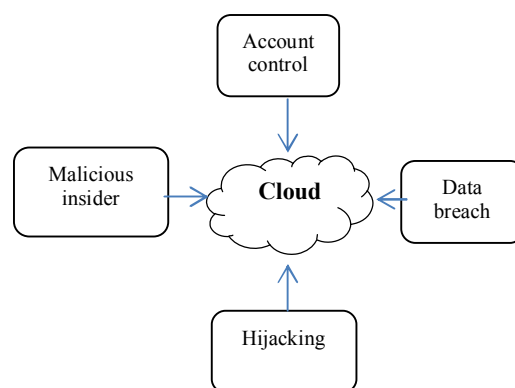

Fig 1: Category of Threats

2.4 COMPLIANCE

For compliance the access will be monitored and track to ensure that there will be no security breaches in the system. compliance requirement become very strict as year are going on and cloud service provider can meet these requirements. Compliance requirements are defined by regulatory bodies, government, for different type of stored data.

Data privacy and business continuity are two big items for compliance, what procedure is put in place for implementing. The cloud security alliance states that the SLA between the cloud customer and the provider should include a right to audit clause, which address audit right

are required to ensure compliance with regulations.

2.5 TRUSTED THIRD PARTY

The term trusted third party in cloud is relate the establishment of trust level and provide a solution to maintain the confidentiality, integrity and authenticity of user's data. A trusted third party is used to facilitates secure interaction between two parties who has trust on each other. And data is secured in end to end services.

Third party review all critical transaction communication between the parties. Third party operationally connected to public key infrastructure (PKI). PKI provide technically legal acceptance means to implement:

- Provide strong authentication
- It provides authenticated access of resource and database.
- Provide proper protection of local or global data.
- We must ensure that no part of electronic transaction can deny of its presence

3. CONCLUSION

Cloud computing having many advantages & must to keep pace with increasing data users inrecent scenario its security issues must be understood. Cloud computing is known for its lowercost, faster and easiest services. But people have different point of view. and some feel cloud is unsafe. So security is main concern in cloud because when we put data on cloud it must be safe. When data is deployed in cloud, security must be deployed and integrity, control, audit and availability will be maintained.In future cloud computing must maintain security & privacy at its full extent as users are increasing day by day.

4. REFERENCES

[1] Armbrust, Michael,et al . "A view of cloud computing" communication of the ACM, 53(4), pp.50-58,2010

[2] Information security magazine, the three cloud computing risk to consider issue: june 2009 retrieved from ttp://www.arma.org/press/ARMAnews/infosecurity.pdf

[3] Huaglory Tianfield " Security issues in cloud computing" , ieee ,pp1082-1089,2012

[4] W Baker"M" 2011 DBIR [online] available www.wired .com/images/threatlevel/2011/04/DBIR.pdf

[5] Wei-tek Tsai, Xin sun, Balasooriya "Service oriented Cloud computing Architecture" ieee, pp.684-689.2010

[6] Zetta," Zetta :enterprises cloud storage on demand", http://www/zetta.net/2006

[7] M. H. Talebian, E. Ahmed, A. Gani, and M. K. Khan, "Areview on remote data auditing in single cloud server: Taxonomy andopen issues," Journal of Network and Computer Applications, vol. 43, pp. 121–141, 2014.

[8] X. Wang, B. Wang, and J. Huang, "Cloud computing and its keytechniques," in Computer Science and Automation Engineering (CSAE),2011 IEEE International Conference on, vol. 2. IEEE, 2011, pp. 404–410.

[9] Cloud security alliance, security guidance for critical areas of focus in cloud computing. V2.1,2009

[10] j.Snyder, six strategies for defense in depth: securing the network from the inside out http://www.arubanetwork.com/pdf/technology/whitepaper/wp_defense_in_depth.pdf

[11] K. Yang and X. Jia, "An efficient and secure dynamic auditing protocol for data storage in cloud computing," Parallel and Distributed Systems, IEEE Transactions on, vol. 24, no. 9, pp. 1717–1726, 2013.]

[12] Z. Qiao, S. Liang, S. Davis, and H. Jiang, "Survey of attribute based encryption," in Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2014 15th IEEE/ACIS International conference on, June 2014, pp. 1–6.

[13] M. Bahrami and M. Singhal, "A light-weight permutation based method for data privacy in mobile cloud computing," in Mobile Cloud Computing, Services, and Engineering (MobileCloud), 2015 3rd IEEE International Conference on, March 2015, pp. 189–198.

❑❑❑

CRUCIAL FACTORS AFFECTING ADOPTION OF CLOUD COMPUTING IN MODERN INDUSTRIES : CHALLENGES AND PERSPECTIVES

Dr. Neeta Sharma

Assistant Professor, CSE Departemnt, Noida International University, G.Noida, Uttar Pradesh, INDIA.
Corresponding Email ID: neeta.msit@gmail.com

ABSTRACT

As Cloud Computing is an advanced technology with better functionality offered internet services on rent base with many advantages such as scalability, flexibility, integration, and cost reduction. In addition, it provides an advanced virtual storage for organizations to deploy their applications to perform various operations. Despite many compelling benefits, there are some crucial factors which are affecting the adoption of Cloud Computing by the different organizations. This is one of the main challenges that encumber the growth of Cloud Computing adversely affecting Cloud Service Providers (CSP). On the other hand, it is heartening to observe cloud service providers' endeavor to migrate all the data/application over the clouds and increase their reliability between CSP and cloud users. In this paper, we have tried to investigate the challenges and different perspectives of adoption of cloud computing in modern industries.

Keywords: Cloud Computing, Security, Availability, Awarenes.s

1.   INTRODUCTION

The evolution and expansion of electronic services requires continuous enhancement in the term of infrastructure. In this context, Cloud Computing offers a comparatively cost effective (pay as per usage) scalable alternative to wherever to use infrastructure. Cloud Computing refers to hosted services online. These services are access via the internet which is symbolically depicted as a 'cloud' and a model which demonstrates a promising future for the cloud users.

Cloud Computing offers various services on different layers such as SaaS (Software as a Service is also known as Software layer and Gmail is best example of this), PaaS (Platform as a Service is responsible for IT services such as Operating System or an application framework and customer also installed or developed additional software) and IaaS (Platform as a Service is responsible for IT services like Operating System or an application framework and customer also installed or developed additional software. With service layers Cloud Computing also provides deployment models. Deployment models are private cloud (a type of cloud where parts of computing resources can also be reserved and dedicated for one organization/customer only is known as private cloud computing), public cloud (a type of cloud where parts of computing resources can be shared to the public), community cloud (a type of cloud where parts of computing resources can be shared infrastructure for specific community and hybrid cloud (A type of cloud where parts of computing resources can also be shared for by the two and more organizations). On the conclusion we can say that Cloud Computing is delivers all types of hardware and softwares [1]. Cloud Computing should assure for good performance, availability and secure storage and transmission.

Conversely, many of the survey and reports shows that there are many factors that organizations are still unresponsive to adopt Cloud Computing such as less awareness, security and privacy of data, multi tenancy virtualization, application security, identity management, access control and authentication etc. and IDC survey report shows that security is the utmost concern of cloud users [2,3,4,5]. So, there is highly requiring serious action to covers their week points. This research paper present a conceptual study of the awareness and security issues related to data and network of Cloud Computing.

2.   LITERATURE REVIEW

In the paper "Cloud Computing Awareness and Adoption Among Small and Medium Scale Business (SMB) in Nigeria (June 2015)", Omotunde A. A. et al opined that Cloud Computing is having its moment in the sun, as the concept of utilizing computing as an on-demand subscription creates operating and economic efficiencies, but Cloud Services Providers and other recognized bodies of Information technology (IT) have to create substantial awareness to utilize the benefits of Cloud Computing [3].

In the paper "Study of Security Issues in Cloud Computing(June 2015)", Versha et al discusses the tenets of this new and emerging technology cloud computing that provides shared resources and services at decreased cost of hardware and software along with some associated security issues while using services of cloud. Further, authors emphasis on the feature of multi tenancy while also discussing the security issues of data on the cloud discovered by CSA. Authors conclude that security issue in cloud computing can be decreased by modifying or designing a strong and robust architecture for multi tenancy [4].

In the paper "A Survey of Cloud Authentication Attacks and Solution Approaches (October 2014)", B. Sumitra et al. have analysed that, it is an evolving computing model that offers great potential, but providers of this fast growing technology require to address many issues such as virtualization, application security, identity management, access control and authentication. This paper mainly focuses on probably all types of authentication attacks that restrict legitimate users' access to services and propose possible mitigation countermeasures in the cloud. Discussing about various authentication attacks such as eavesdropping, man-in-the-middle, cookie poisoning, replay, Session hijacking, malware injection, DoS, insider attacks etc. the authors suggest that user authentication mechanism should be strong enough to protect cloud from various types of attacks [5].

In the paper "A Survey on Security Issues and Solutions At different layers of Cloud Computing (Feb 2013)" Chiraj Modi et. al. have opined that, since cloud services are delivered using classical network protocols and formats over the Internet, implicit vulnerabilities existing in these protocols as well as threats introduced by newer architectures raise many security and privacy concerns. The factors affecting Cloud Computing adoption, vulnerabilities and attacks, and identify relevant solution directives to strengthen security and privacy in the cloud environment [6].

In the paper "An overview and Study of Security Issues & Challenges in Cloud Computing (September 2012)" Rajesh Piplode el discovered through survey that the major obstacle for adoption of cloud is concern for security. They list out ten security concerns in details and some solutions for security issues in Cloud Computing including Investigation Support, Network Security, Encryption Algorithm, Backup, and Customer satisfaction. Most importantly, the authors have proposed a Security Management Model called CMM which describes twenty recommended security management models and their requirements for cloud computing that cloud service providers should definitely consider as they develop or refine their compliance programs [7].

In the paper "A Brief Survey on Architecture, Challenges and Security Benefit in Cloud Computing (February 2012)" Ajith Singh .N el have expressed their worries about bigger companies that still lack the conviction that their bulk data will be maintained with the utmost privacy over the internet. This concern can only be resolved by drafting strict rules for maintaining data privacy over Cloud Computing and strict adherence to these rules. Invention of internet changed the way we use of computer and cloud is going to change that to other level [8].

3. MOTIVATION

As per litterer review we found that Cloud Computing provides many advantages like scalability, cost-reduction, integration but still many organizations / industries / consumers are reluctant to adopt Cloud Computing services. To investigate the reason we carried out this internet based survey study as it has the following advantages:

- *Easy:* It is easy or convenient method to design and conduct the study with the help of Google forms. It can also quantify our recorded data.
- *Cost effective:* There is no cost occur during this study during the design and circulate the survey/tool.
- *Select audience:* On the base of convenience sampling (a type of data collection method) we can easily choose the target audience with the help social networking site and email addresses.

4. METHODOLOGY

To achieve this objective, we have designed a questionnaire for primary data collection based on convenient sampling into three segments such as IT industry, Education Industry and individual consumers from different organizations.

4.1 SOURCES DATA COLLECTION

Secondary Data: Data collection from the printed or published source such as books, journal, research papers, articles and newspapers falls under Secondary data. It is the easy way of data collection. Secondary data analysis also saves the time, provides larger and high quality databases that would be not be easy for any individual researcher to collect on his/her own.

Primary Data: Primary data are directly collected from the respondents with the help of questionnaire. For primary data collection, we have used an online questionnaire for which there is a web based survey tool called "FACTORS AFFECTING ADOPTION OF CLOUD COMPUTING IN MODERN INDUSTRIES". We prepared the questionnaire keeping in mind our objectives. It is cost effective, less time consuming, easy to use and easy to analyze the results. There is no limit for number of questions. To analyze our results as per our requirement, firstly we quantify the data and generate Pie charts and Bar Charts.

4.2 DESIGNING QUESTIONNAIRE

Designing questionnaire is a most important task. We

prepared questionnaire on the basis of the conclusion in the literature review in order to achieve our objectives. We prepared the questionnaire in such a way as to wrap up our research questions in a precise manner. After designing a questionnaire, we have made a pilot study of 30 respondents to check the results. After compilation of whole result of pilot study we finalized our questionnaire for a full fledged survey.

4.3   TARGETED POPULATION

After a literature review, on the basis of our objectives, we have arrived at the decision to target small and middle scale IT industries, Education industries, some consumers from different non-IT industries and also those who are directly and indirectly connected with internet under the categories of 'others'. Within these industries we specially targeted IT personnel, network administrators, software developers, faculty, directors, analysts, server administrators and individual consumers etc. for this survey. While cloud computing is in nascent stage and it is not so common among the various companies and consumers, so to find out the right people who are conversant with cloud computing was a difficult task.

4.4   RESEARCH OBJECTIVES

This survey questionnaire has been designed with the following objectives:

- To evaluate the industry specific response on awareness about Cloud Computing
- To explore the industry specific response for the willingness to use/adopt cloud computing for (parts of) IT.
- To analyze crucial factors affecting adoption Cloud Computing.

5.   SURVEY RESULTS

As this survey is mainly focused on evaluating the factors that affect the adoption of cloud computing in the organizations. Our survey questionnaire was designed to investigate the various factors impact the adoption of cloud computing, such as awareness, security threat, privacy, cost, Insecure availabiltity. Some of the survey results using graphs and charts is as follows:

Q1. What does Cloud Computing primarily mean for your organization (As per the IT, Education and Other industries).

As conclustion of all data collection by the different industries (IT and Education) and individual consumers from different organisations.

Fig. 1: Cloud Computing Understanding

Q2. What are your main concerns regarding the use of cloud computing?

Fig. 2: Concern regarding use of cloud computing

Q3. What are your main Concerns related to the Security & Privacy in the Cloud Computing?

Fig. 3: Concern regarding Security & Threats on cloud computing

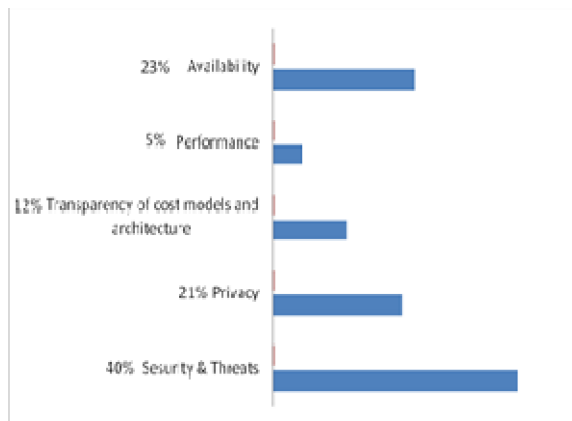*Q4. What aspects of cloud computing should be improved?*

Fig. 4: Improved aspects of cloud computing

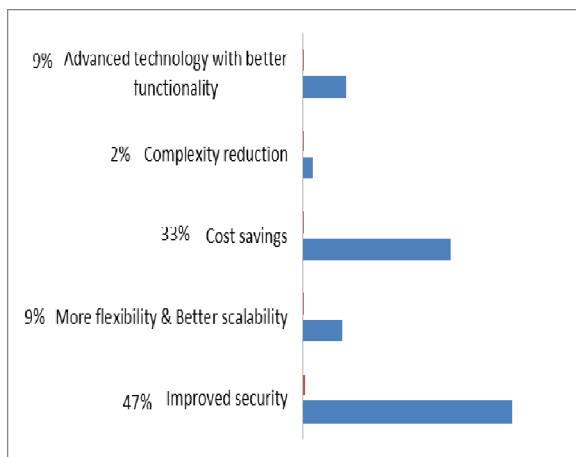*Q5. What benefits do you expect from cloud computing?*

Fig. 5: Expected benefits from cloud computing

6. RESULT INTERPRETATION & CORRELATION ANALYSIS
6.1 RESULT INTERPRETAION

The interpretation of the survey results indicates that while the awareness level about cloud computing is satisfactory among most of the IT industry, consumers, there is a lack of awareness about cloud computing technology among the employees of Education industry and other Non-IT industries. There is no clear understanding about the capabilities and tenets of Cloud Computing among the employees of Education industry and other Non-IT industries which is not a healthy sign for future proliferation of this powerful technology in general and the business proposions for cloud service providers in particular.

Through the responses to item no 2 &3, an inference could be drawn about the reseans for not adopting cloud computing. It was revealed that the single most important factor resisting the adoption is concern for security issues with 42% strongly agreeing and more than 70% agreeing to this statement.

Item number 4 was posed to furhter investigate on what item 3 and 5 had revealed. In other words, under securty concerns whether data security was most prominent concern or the network security. Nearly 40% consumers showed Data security as their most prominent concern whereas network security was percieved as most prominent concern by 27 percent of the respondents. There were some other concerns alsoas depicted in figure 22 that contributed to adoption of cloud how ever these remain less significant.

6.2 CORRELATION ANALYSIS

Correlation Analysis Correlation indicates the direction of co-variance and denoted by 'r'. The value of 'r' lies between ± 1. Positive values of r indicate positive correlation between the two variables (i.e., changes in both variables take place in the statement direction), whereas negative values of 'r' indicate negative correlation i.e., changes in the two variables taking place in the opposite directions. A zero value of 'r' indicates that there is no association between the two variables. When r = (+) 1, it indicates perfect positive correlation and when it is (−) 1, it indicates perfect negative correlation, meaning thereby that variations in the independent variable (X) explain 100% of the variations in the dependent variable (Y). We can also say that for a unit change in the independent variable, if there happens to be a constant change in the dependent variable in the same direction, then the correlation will be termed as perfect positive. But if such change occurs in the opposite direction, the correlation will be termed as perfect negative. The value of 'r' nearer to +1 or −1 indicates a high degree of correlation between the two variables.

Correlation Table

Variables	Meaning of cloud	Adoption of cloud	Cost & legal Concern	S&T Blocking	Aspect improve.	S_P Issue	Expectation
Meaning of cloud		.435**	.251**	.188*	.271**	.118	.247**
Adoption of cloud			.145	.071	.335**	.051	.066
Cost & legal Concern				.173*	.382**	.119	.501**
S&T Blocking					.171*	-.058	.027
Aspect improvement						.202**	.439**
S_P Issues							-.030
Expectation							

Fig. 6: Correlation Table

7. CHALLENGES & PERSPECTIVES

The survey questionnaire consisted questions on cloud awareness and cloud security related issues. Some useful insights that have emerged as the interpretation of the

survey results, could be put to good use by both the cloud service providers and industries to enhance adoption of cloud computing in the future. As well as there is a need to conduct awareness programs, in-house training programs, improve confidentiality, integrity and availability, improve network security through collaboration, improvement of web-security, improvement of authentication and authorization process to increase the adoption of cloud computing.

8.   CONCLUSION

An online survey has been conducted to investigate the factors affecting cloud adoption by different industries such as; IT industry, Education industry and non-IT industries. Our result interpretation has revealed that there is a large gap in understanding the cloud computing technology among consumers. Specially the majority of the education industries and non-IT industries are unclear and unknown about the cloud computing technology. The majority of consumers perceive that cloud computing is not secure and security is the main blocking factor to limit the adoption of cloud services. The network security threat is of immense concern, followed by data security concerns. Lower cost is only positive consideration for implementation of cloud services in organizations.

9.   REFERENCES

[1] Osama Harfoushi, Bader Alfawwaz, Nazeeh A. Ghatasheh, Ruba Obiedat, Mua'ad M. Abu Faraj, and Hossam Faris, "Data Security Issues and Challenges in Cloud Computing: A Conceptual Analysis and Review" Scientific Research, Communications *and Network*, Published Online February 2014.   http://www.scirp.org/journal/cn,

http://dx.doi.org/10.4236/cn.2014.61003, 2014, 6, 15-21.

[2] F. Gens, "New IDC It Cloud Services Survey: Top Benefitsand   Challenges," 2009.http://blogs.idc.com/ie/?p=730

[3] Omotunde A.A. Izang, A.A., Awonyi O.C., Omotunde, B.K, and Mensah, Yaw.A, "Cloud Computing Awareness and Adoption Among Small and Medium Scale Business (SMB) in Nigeria (June 2015)", International Journal of Multidisciplinary Sciences and Engineering, ISSN:2045-7057, Volume 6, Number 6, Page 32-38, June-2015.

[4] Varsha, Amit Wadhwa and Swati Gupta, "Study of Security Issues in Cloud Computing", International Journal of Computer Science and Mobile Computing (IJCSMC), ISSN: 2320-088X, Volume 4, Issue 6, Page 230-234, June-2015, www.ijcsmc.com.

[5] B. Sumitra, C. R. Pethuru, and M. Misbahuddin, "A Survey of Cloud Authentication Attacks and Solution Approaches", International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), An ISO 3297:2007, ISSN (online): 2320-9801, ISSN (Print): 2320-9798, Volume 2, Issue 10, October 2014, www.ijircce.com.

[6] Chiraj Modi, Dhiren Patel, Bhavesh borisaniya, Avi Patel, and Muttukrishnan Rajarajan, "A Survey On Security Issues and Solutions At different layers of Cloud Computing", The Journal of Supercomputing, Volume 63 Issue 2, February 2013.

[7] Rajesh Piplode, and Umesh Kumar Singh, "An overview and Study of Security Issues & Challenges in Cloud Computing", International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 2, Issue 9, September 2012.

[8] Ajith Singh. N, Vasanthi. V and M. Hemalatha, "A Brief Survey on Architecture, Challenges & Security Benefit in Cloud Computing", International Journal of Information and Communication Technology Research, Volume 2 No 2, Page No-102 and 109, February 2012.

❑❑❑

CLOUD COMPUTING SECURITY FUNDAMENTALS

Pooja Sharma

Assistant Professor, Department of Computer Science & Engineering,
Accurate Institute of Management and Technology, G.Noida, Uttar Pradesh, INDIA.
Corresponding Email ID: pooja.aquarius12@gmail.com

ABSTRACT

As cloud is gaining more popularity, more and more organizations are wanting to move towards the cloud. But the key concern about moving towards cloud would has been security. So chief information officer of an organization while deciding to move to the cloud, he would have lot of questions. Is my data Secure on Cloud? Can others access my confidential data? For example, if a competitor also using same cloud infrastructure, how safe is my data, how confidential is my data. Also there are lot of government regulations and compliance perspective. So how can I make sure that my infrastructure, my cloud infrastructure is compliant with the government regulations and what if an attacker brings down my app which is hosted on cloud. How do I avoid this problem and how do I remediate if such an attack is to happen. So these are the major security concerns which are stopping the organizations from moving completely to cloud. So in this paper, I will walk you through information security concepts. So how cloud security is different from traditional security . How the cloud infrastructure landscape is and how it is different from the traditional classical data center. Concerns and threats in cloud infrastructure and the security mechanisms which will be enforced in cloud are discussed.

Keywords : Cloud Computing; Security; CIA; Virtual Machine (VM); Cloud Service Provider (CSP); Operating System (OS); IDS.

1. INTRODUCTION

Cloud computing is today's one of the most popular technologies as it has the ability to reduce cost associated with computing while increasing flexibility and scalability for the computer processes. Within these past few years, cloud computing has brought various new business ideas to one of the most successful parts of IT industry[1]. It is helpful in the fast growing of IT industry. There are various critical concerns which are expressed by IT organizations. One of the major concerns is security. Security problem has more exaggerated under the cloud model as various new problems has entered in cloud computing related to multi-tenancy, layered architecture, etc. In this paper, the security concerns are discussed, providing the security threats and security mechanisms/ measures.

2. INFORMATION SECURITY- CIA TRAIDS

Confidentiality - Confidentiality of data ensures that your data is confidential. Any unauthorized user cannot access your data. Only authorized user can access your data.

Integrity: Integrity makes sure that your data is remains as it is. So no unauthorized user can change your data. Making sure that your data is not medalled with [1].

Availability: Availability makes sure that your data and services and applications are always available to authorized users.

So CIA triads are a key security concept. All our security mechanisms revolve around these three security triads. Similarly three A's (AAA) are another popular concept in identification and authentication management.

Fig. 1: CIA Traid

Authentication – It is the method of verifying who the user claims to be. For example user says I am user A, How will one ensure that user is really user A. So for this , one of the simplest mechanism is password. Only password may not be sufficient, further in paper additional authentication mechanisms which are available are presented.

Authorization - It is the method for ensuring that only authorized person can get access to the data. For example,

the finance department of a company might have access to only the finance data. So an HR person who is trying to access the finance data should not be given access if the user doesn't have access to data.

Auditing – It is a process of going through all the activities of system in your infrastructure and keeping track of what has happened. In case of an attack and in case of something which is anomaly, so going back and being able to check what has happened in your system. So auditing is one of the main compliance of the requirement.

2.1 MULTI-FACTOR AUTHENTICATION

As mentioned above, using password for authentication is not sufficient. Nowadays a lot of password cracker software's are available and may be some people use very simple passwords like their kids name, or birth dates , etc. or something like that which is easier to guess. So password being only one factor of authentication is not very secure. In order to avoid that , we need to add additional layers of security. So the concept of multi-factor authentication is that in addition to password, you will have multiple layers of security. One of the examples of multi-factor authentication is using OTP (One Time Password). So there are physical tokens available as well as software tokens available which generate OTP. So when user is trying to authenticate, user gives the user name, password, which a person knows and key which the person physically has. So this has additional layer of authentication. Also there are concept of bio-metric authentication, where we have finger print recognition and retina scanning.[2] So highly confidential data uses biometric authentication which is much more secure than any other forms of authentication.

Encryption is commonly used for securing your data. So it's a process of converting data in a form that can be used in a meaning manner only if one has a specific knowledge. So in case of symmetric key and asymmetric key or public key cryptography, the user will have a public key and a private key. So it is very specific to that user, only if you know what key is , then you can access the data otherwise the encrypted data for the another user is meaningless. So the process of changing the encrypted data back to the original data is called decryption.

2.2 DEFENCE-IN-DEPTH

Defense in depth is one more mechanism in security where you have multiple layers of security. So if you have only anti-virus in your desktop, some attack through the network may not be captured. So you have multiple layers. Perimeter Security which is physical security. One make sores that there server are in a safe location[2]. An unauthorized user can come and steal your hard-disk, so it is important to have physical security in the first place. Usually you have access cards or other methods to make sure that physical security is ensured. Network Level Security where there is firewall and DMZ Compute Security is anti-virus and data loss and prevention

products. Storage Security is encryption and zoning. If in one layer of security, if a attacker gets to your infrastructure, you make sure that in the another layer at least you detect this attack and stop this attack. Hence this will improve the security of the system instead of having only one mechanism of security.

2.3 SECURITY- TRADITIONAL V/S CLOUD

The major difference between traditional data center and modern data center which is hosted on cloud. In traditional data center all your resources like server, storage and everything might be co-located in a single location or may be in multiple location but is limited to that organization. In case of private cloud is, it might be restricted to only the organization but in case of public cloud or hybrid cloud, the infrastructure is shared by multiple organizations and is used by multiple users and it might be very huge, distributed across geography[3]. The one server might be in Asia and other server might be in US. One might be using it seamlessly without knowing where your server is located and where your data exists[3]. You have mobile apps which are trying to access cloud, so it is more complicated than a traditional data center and you need in depth security mechanisms to make sure that your cloud infrastructure is secure.

2.4 CLOUD DEPLOYMENT MODEL

Public Cloud: Suppose there is an enterprise P and enterprise Q and user might belong to either of the organizations. Public cloud is hosted by CSP (Cloud Service Provider). Service provider is responsible for maintaining your server, storage, network and all the resources and making sure that when requested, the resources are available to you. But the user need not know what is there at the back-end. Also it is shared by different organizations, different users. So user is not aware of who also sharing his data.

Private Cloud: In case of private cloud, it is restricted to a particular organization. It might be on premise (in the enterprise) or it can be hosted by a cloud service provider but the entire infrastructure is restricted to one enterprise. There are different multiple enterprises using the same infrastructure.

Hybrid Cloud: Hybrid cloud is example a company where a mission critical data is in its private cloud , in its enterprise cloud where it make sure that my mission critical very confidential data is not on cloud but some not so important data might have moved to public cloud. So it's a combination of private and public cloud which is hybrid cloud[4].

3. CLOUD SECURITY CONCERNS

So with the introduction of multiple users and multiple enterprises and a heterogeneous hardware infrastructure, new security concerns arises.
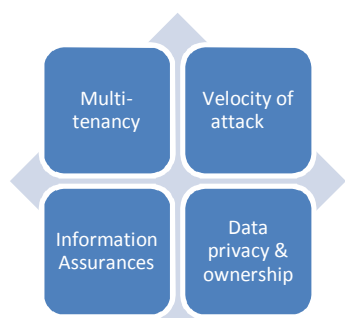
Fig. 2: Cloud Security Concerns

3.1 MULTI-TENANCY

It is a concept where multiple virtual machines are located in a single infrastructure, single server or a single server host. Each virtual machine is co-located in a same server. So in case of a public cloud, Example: My Company might be using one virtual machine, and another company the competitor company might be using another virtual machine which is hosted on the server. So that's how multi-tenancy brings in a security threat, where the same infrastructure might be shared in the different organization and your virtual machines might be co-located in a single server. So for the CSP also it brings new challenge[5]. Each company has its own security policies, so when multiple organizations have heterogeneous various forms of security policies, how does the cloud provider ensures that they each company security policy is fulfilled or is enforced very correctly in a uniform way because the underlined infrastructure is the same.

3.2 VELOCITY OF ATTACK

In a public cloud might be thousands of server running in the same location or across the geography . So the entire infrastructure available to user, hence surface of attack increases with the introduction of cloud. In case of single enterprise with the single data center which is within its premise, its only there set of server only which they have control. in cloud the infrastructure is huge so the surface of attack is huge, that's why the velocity of attack is higher. Because of this potential loss also is very high because if one VM is attacked, the entire infrastructure might get attacked. So in that case, trying to contain the attack to the particular location is also become more difficult. To counter this challenge, the CSPs need to have more strict security enforcement mechanisms compared to a classical data center.

3.3 INFORMATION ASSURANCE AND DATA OWNERSHIP

How do you ensure that confidentiality of your data is maintained, because CSP has now access to your data and also some other enterprise might be using the same infrastructure? How do you make sure that your competitor is not stealing your data?.[4] They major

security concerns are CIA (Confidentiality, Integrity and Availability). In case of cloud environment, your data is hosted by CSP. So the CSP has access to data, but the owner is not the CSP. The organization is the owner. How to make sure that your data is accessed only by the authorized user and ensuring that the confidentiality is maintained. So the data also should be protected by the unauthorized use through encryption. So this becomes one of the major concerns in cloud environment.

3.4 DATA PRIVACY

How one makes sure that privacy of data is ensured in the cloud environment?

Potential of unauthorized disclosure of private data of a cloud client. Private data may include:
- Individual identity of the client
- Details of the service requested by the client
- Proprietary data of the client

A CSP needs to ensure that private data of its client is protected from unauthorized disclosure.
- Both collection and dissemination of private data requires protection
- A CSP needs to deploy data privacy mechanisms, which are complaint with the regional legal regulations.

4. CLOUD SECURITY THREATS
4.1 VM THEFT

Virtual Theft is a vulnerability that enables an attacker to copy or move a VM in an unauthorized manner. VM Theft is a result of inadequate controls on VM files allowing unauthorized copies or move operations. Copy and move restrictions are essential to safeguard against VM theft.
- These thefts bind a VM to a to a specific physical machine.
- A VM with a copy and move restriction cannot run on a hypervisor installed on any other server.
- These restrictions' uses a combination of virtualization management.
- Limit applying such restrictions to critical / sensitive VMs only.
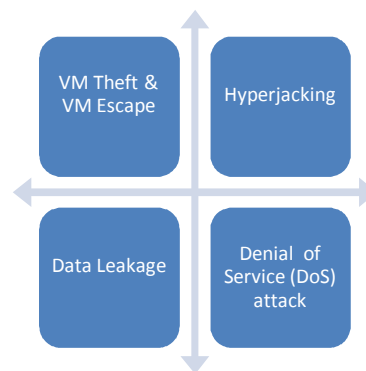

Fig. 3: Cloud Security Threats

4.2   HYPERJACKING

It enables an attacker to install a rogue hypervisor or Virtual Machine Monitor (VMM) that can take control of the underlying server resources. An attacker can run unauthorized applications on a guest OS realizing it. An attacker could control the interaction between the VMs and the underlying server[5]. Regular Security measures are ineffective against hyper jacking. Measure against hyper jacking includes:

- Hardware assisted secure launching of the hypervisor
- Scanning hardware level details to assess the integrity of the hypervisor and locating the presence of rogue hypervisor.

4.3   DATA LEAKAGE

Confidential data stored on a third party cloud is potentially vulnerable to unauthorized access or manipulation.

- Attacks on service providers controls system (for example password lists) could make all the clients data vulnerable.
- Cloud users must evaluate end-to-end data protection measures by all the concerned parties who have any level of access on the data.

Side Channel Attacks (SCA) can be used for data leakage in cloud:

- An SCA extracts information by monitoring indirect activities for example cache data.
- Cross-VM SCA cloud reveal information of a cloud to another malicious client that runs its VMs on the same server.
- Protection against cross VM SCA requires placing only those clients that have no conflicts with one another on the same server.

4.4   DENIAL OF SERVICE (DOS) ATTACK

It is an attempt to prevent legitimate users from accessing a resource or service[6]. DoS attack might affect software applications and network components. A DoS attack involves:

- Exhausting resources, for example network bandwidth or CPU Cycle
- Exploiting weaknesses in communication protocols, for example resetting of TCP sessions, corrupting domain name servers' cache.
- A malicious client VM might be used to launch DoS attack against the hypervisor or other VMs running on the same hypervisor.
- As a protective measure, resource consumption of a VM needs to be restricted.

5.   CLOUD SECURITY MECHANISMS
5.1   COMPUTE A NETWORK LEVEL SECURITY

In a Virtual environment, the compute, network and storage all are virtualized. So at that level, how one ensure

that security is maintained. Securing a computer system includes

- Securing Physical Server
- Securing hypervisor
- Securing VMs(VM isolation and VM hardening)
- Security at guest OS level (Guest OS Hardening)
- Security at application Level (Application hardening)

Each VMs have to be secured because if one VM is compromised, it can act as an attack point for the rest of the VMs.

PHYSICAL SERVER SECURITY: Identifying physical server application details include :

- Whether server will be used for specific application or for general purpose
- The network services provided on the server
- Users and/or user groups who can operate the server and their access privilege

Deciding protection measures:

- Determining authentication and authorization mechanisms
- Disabling unused hardware such as NICs, USB ports, or Drives.

HYPERVISOR SECURITY: attacks on the hypervisor impact all the VMs running on it. We can say, it is also a single point of failure. Hypervisor Security measures are

- Install hypervisor updates
- Harden VMs to prevent attacks

Protection of the hypervisor management system

- It is critical because an insecure management system can make existing VMs vulnerable for attacks and enable creation of new malicious attacks
- Configuring strong security on the firewall between the management system and the networks
- Providing direct access only to administrators to management servers
- Disable access to the management console to prevent unauthorized access.

VM SECURITY: ISOLATION AND HARDENING: VM isolation helps prevent a compromised guest OS and application running on it from impacting other VMs . VM hardening is a process of changing the default configuration in order to achieve greater security. The considerations of VM Hardening are

- Use VM template to to provision new VMs
- Limit the resources that VM can consume to prevent DoS attacks
- Disable unused functions and devices on VM
- Use a directory service for authentication
- Perform vulnerability scanning and penetration testing of the guest OS.

GUEST OS AND APPLICATION SECURITY : Guest OS hardening mechanism includes deleting unused files and applying the latex patches[6]. Applying hardlist

checklist available for specific Oss. Installing the guest OS in TCB mode if the VM is to be used for critical applications. It needs a support from hypervisor in configuring (trusted) virtual hardware components for TCB. Application hardening measure includes disallowing a vulnerable application from

- Launching any untrusted (executable) file
- Creating or modifying executable files
- Modifying sensitive areas of the guest OS , for example MS Windows registry

Sandboxing is another important measure for guest OS and application security.

SECURITY AT NETWORK LEVEL - VIRTUAL FIREWALL : Securing VM-to-VM traffic running on a server is difficult in a VDC environment.

- Virtual switches could be invisible to administrators (network and systems)
- Traffic may never leave the server, so it cannot be detected and intercepted

Virtual firewall is a firewall service running on the hypervisor.

SECURITY AT NETWORK LEVEL : It is a physical or logical network or sub network that limits the exposure of the nodes in the internal network from the external networks. It adds additional layer of security against external attacks.

- An attacker has access only to the DMZ, rather than any other part of the network
- For practical purposes, services provided to the users o the external network can be placed in the DMZ.

A virtualized DMZ is a DMZ established in a virtualized environment using virtual network components.

5.2   SECURITY DATA AT REST

Data at rest is the data which is stored in the server or in any of the storage locations . If the data is not encrypted or if it is the raw data which is stored in your physical storage. If an attacker gets access to your data, it become easier for the attacker to make use of that data[7]. So if you encrypt your data which is at rest, even if the attacker gets hold to your data, it is in encrypted form, so it is of no use to the attacker. So it is very important to secure your data at rest which is done through encryption.

Data at rest means data which is not being transferred over the network. Encryption of data at rest is done by:

- Providing confidentiality and integrity services
- Reduces legal Liabilities of a CSP due to an unauthorized disclosure of data on its cloud.

Full disk encryption is a method to encrypt data at rest residing on a disk.

SECURITY AT NETWORK LEVEL: It is securing data in flight. Data will be at rest like files which is stored on server. But there will be a data which is being transferred over network i.e moving where your files are hosted on a location and your client is trying to access the file. At network level also, data can be stolen and compromised, so one need to have network level security to ensure that data which is in flight is also secured. Encryption of data in flight provides confidentiality and integrity and also it is a key measure against sniffing attacks.

Table 1: Security at Network Level

Encryption Method	Description/ Example
Application Level	Applied at the application level where data is generated
Network Level	Applied at the network level; for example, IPSec to encrypt IP packets

Data Shredding: It is a mechanism where the data is erased from the desk permanently and making sure that no residual data is left. The residual data can be used by the attacker, so it must be avoided.

Data which is deleted by a cloud client or a process , but which leaves traces on the system , can be potential source of attacks.

- Traces of deleted VMs can provide vital information to an attacker
- Partially recoverable "deleted data" may reveal clients details

Data shredding permanently removes all the traces of the deleted data. It is a critical feature for data security in cloud Infrastructure. Traces of the deleted data include

- Logs of VMs or application executions
- Logs of old files, folders and other resources
- Logs of data communication.

INTRUSION DETECTION : It is a process of detecting events and/or entities that could possibly compromise the security of the system.

Table 2: Types of Intrusion Detection

Types of Intrusion Detection System (IDS)	Description
Server based IDS	Analyzes activity logs, including system calls, application logs, etc. Better View of the monitored system but high vulnerability for an attack on IDS itself
Network based IDS	Analyzes network traffic and communicating nodes Poorer view of the system and low vulnerability for an attack on IDS itself
Integrated IDS	Combination of server and network based approaches

5.3 IDENTITY ACCESS AND MANAGEMENT IN CLOUD

One Time Password
- Every new access request requires new password
- A measure against password compromises

Federated identity management is provided as a service on cloud
- Enables organization to authenticate their users of cloud service using the chosen identity provider
- User identities across different organizations can be managed together to enable cloud collaboration on cloud

OpenID
- It is an open standard for decentralized authentication and access control
- Can be used while allowing users to log onto many services using the same digital identity

5.4 RISK ANALYSIS AND COMPLIANCE

Risk refers to the effect of uncertainty on business objectives. Risk management is a co-ordinated activity to direct and control an organization, and to realize business potential while managing negative events[7]. Risk assessment aims to identify the potential risk while operating in a cloud environment. It should be performed before moving to a cloud. It should be used to determine the actual scope for cloud adoption.

Table 3: Types of Compliance

Types of Compliance	Description
Internal policy compliance	• Controls the nature of IT operations within an organization • Needs to maintain same compliance even when operating in cloud
External regulatory compliance	• Includes legal legislations and industry regulations • Controls the nature of IT operations related to flow of data out of an organization • May differ based upon the type of information , business , etc.

Compliance refers to the act of adhering to and demonstrating adherence to external laws and regulations as well corporate policies and procedures. Cloud adoption and operation for enterprise businesses need to abide by compliance policies. Types of compliance are given above in TABLE 2.

6. CONCLUSION AND FUTURE SCOPE

As the companies are moving towards cloud, so one the major reasons why companies are moving towards cloud is the lower operational cost. So running your applications in cloud is very much cheaper than having running it in a classic data center where you have to buy your own hardware and software. But there is a chance of increased risk in moving to cloud. So the ultimate goal of cloud computing is to decrease the operational cost and increase the revenue and also reduce the risk. So cloud security . So cloud security mechanisms and cloud security measures taken by the company ensure that the risk of moving towards cloud has been reduced. A company can make use of cloud computing to decrease their operational cost and increase their revenue. This is the ultimate goal of cloud computing.

7. ACKNOWLEDGMENT

8. REFERENCES

[1] BEHL A. BEHL K. " An analysis of cloud computing security issues" center of Excellence , Cisco Syst. , Information and Communication Technologies (WICT), 2012 world congress on Oct 30 2012 pp. 109-114.

[2] Sabahi, F. Fac. of Comput. Eng., Azad Univ., Shahrekord, Iran, " Cloud Computing Security Threats and Responses" Communication Software and Networks (ICCSN) , 2011 IEEE 3rd International Conference on 27 May 2011, pp. 245-249.

[3] https://www.sans.org/course/cloud-security-fundamentals

[4] International Organization for Standardization (ISO), "ISO/IEC 27000 - Information technology - Security techniques - Information security management systems - Overview and vocabulary," ISO/IEC 27001:2005(E), 2009, <http://webstore.iec.ch/preview/info-isoiec27000%7Bed1.0%7Den.pdf>, Accessed in July 2010.

[5] Cloud Security Alliance Group, "CSA-GRC Stack," <www.cloudsecurityalliance.org/grcstack.html>, Accessed Dec'10

[6] M. Almorsy, J. Grundy, I. Mueller, "An analysis of the cloud computing security problem," In the proc. of the 2010 Asia Pacific Cloud Workshop, Colocated with APSEC2010, Australia, 2010.

[7] R. La'Quata Sumter, "Cloud Computing: Security Risk Classification", ACMSE 2010, Oxford, USA.

❑❑❑

DESIGN AND SIMULATION OF HIGH LEVEL LOW POWER 7T SRAM CELL USING VARIOUS PROCESS & CIRCUIT TECHNIQUE

Sachin Dubey
Assistant Professor, Department of Electronics & Communication Engineering,
Accurate Institute of Management & Technology, Greater Noida, Uttar Pradesh, INDIA.
Corresponding Email ID: sachindubey.mail@gmail.com

ABSTRACT

Low power memory is required today most priority with also high stability. The power is the most important factor for today's technology so the power reduction for one cell is vital role in memory design techniques. In this paper, we introduced some design circuit techniques for low power design. Leakage current in standby mode is the major part of power loss. We concentrate on the technique that to reduce the leakage current in standby mode.

Keywords: CMOS, SRAM, Threshold Voltage, Circuit techniques, Process Technique.

1. INTRODUCTION

Static random access memory (SRAM) is a type of volatile semiconductor memory to store binary logic '1' and '0' bits. SRAM uses bistable latching circuitry made of Transistors MOSFETS to store each bit. When the cell is selected, the value to be written is stored in the cross-coupled flip-flops. A basic SRAM cell consists of two cross coupled inverters forming a simple latch as storage elements and two switches connecting these two inverters to complementary bit lines to communicate with the outside of the cell shown in fig-1.

Fig. 1: Cross-coupled inverters SRAM Memory Cell

The one CMOS transistor leakage current due to various parameter is the vital role of power consumption. The CMOS leakage current at the process level can be decreased by some implement on deep sub micron method. The circuit level technique is reduced power

Identify applicable sponsor/s here. If no sponsors, delete this text box. (sponsors) consumption at very high level. In this paper, we simulate the 7T SRAM cell using many techniques both circuit level, process level in one cell as Hybrid cell.

2. THE 7T SRAM CELL

The circuit of 7T SRAM cell is made of two CMOS inverters that connected to cross coupled to each other

with additional NMOS Transistor which connected to read line and having two pass NMOS transistors connected to bit lines and bit-lines bar respectively. Fig-2 shows circuit of 7T SRAM Cell, where the access transistors MN3 is connected to the word-line (WL) to perform the access write and MN4 is connected to the Read-line (R) to perform the read operations thought the column bit-lines (BL and BLB). Bit-lines act as I/O nodes carrying the data from SRAM cells to a sense amplifier during read operation, or from write in the memory cells during write operations. All transistors have minimum length (LMIN =45nm according to used Technology), while their widths are typically design parameters. The value of WP1 and WP2 defines PMOS transistors width and WN1 and WN2 defines the NMOS driver transistors width use in CMOS Invertors, while WN3 and WN4 is the access transistors width.
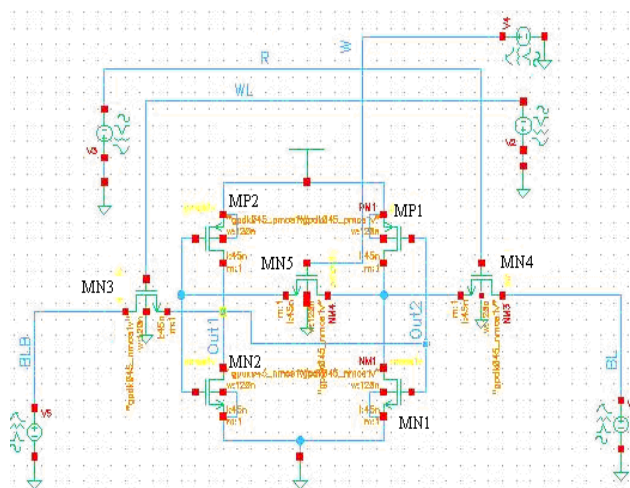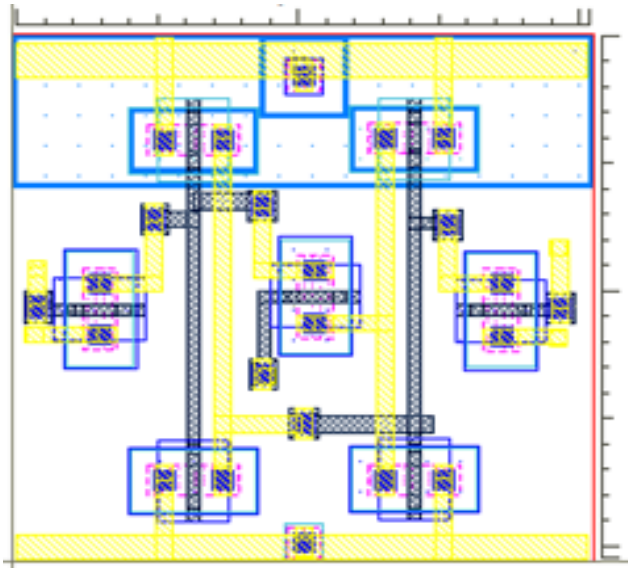

Fig. 2 (a): Schematic

Fig. 2 (b) : Layout
Fig. 2 Seven Transistor SRAM Cell

- Reverse-biased junction leakage current (I_{REV})
- Sub threshold (weak inversion) leakage (I_{SUB})
- Oxide tunneling Current (I_G)
- Gate current due to hot-carrier injection
- Gate induced drain leakage ($IGIDL$)
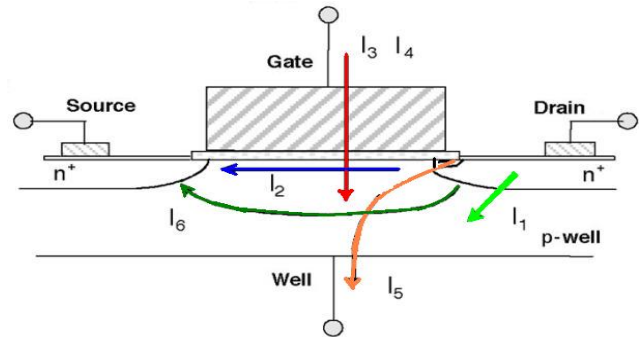- Channel punch through current

Fig. 4: Leakage current mechanism in CMOS Transistor

3.   CMOS LEAKAGE CURRENT

In the CMOS devices, the leakage current is becoming a major contributor to the total power consumption. In current deep-sub nanometer technology with low threshold voltages, sub threshold and gate leakage have become the dominant sources of leakage and are expected to increase with the technology scaling. The leakage power is becoming a significant component of the total power and may contribute to majority of the power dissipation in future *CMOS* technologies [1]. The leakage current and leakage power are increasing with scaling. The two main sources of power dissipation in CMOS circuits are dynamic power dissipation and static power dissipation. Static power dissipation is due to leakage current when the transistor is normally off. The improvement in technology scaling has introduced very large sub threshold leakage current, therefore careful design techniques are very important in order to reduce sub threshold leakage current for low power design. Leakage current occurs in both active and standby modes. It is recommended to switch off the leakage current when the circuit is in standby mode. The power and technology graph shown in fig. 3.

Fig. 3.: Static and Dynamic Power Trends

There are six main sources of leakage current in a CMOS transistor as shown in Fig-4.

3.1   REVERSE-BIASED   LEAKAGE CURRENT

Drain and source to well junctions are typically reverse biased, causing p-n junction leakage current. A reverse-bias p- n junction leakage (I_1) has two main components: one is minority carrier diffusion/drift near the edge of the depletion region; the other is due to electron-hole pair generation in the depletion region of the reverse-biased junction [2]. For an MOS transistor, additional leakage can occur between the drain and well junction from gated diode device action (overlap of the gate to the drain-well pn junctions) or carrier generation in drain to well depletion regions with influence of the gate on these current components [3]. pn junction reverse-bias leakage (I_{REV}) is a function of junction area and doping concentration [4].

3.2   SUB THRESHOLD LEAKAGE (ISUB)

The Sub threshold is the drain-source current of a transistor operating in the weak inversion region. Unlike the strong inversion region in which the drift current dominates, the sub threshold conduction is due to the diffusion current of the minority carriers in the channel for a MOS device. In current CMOS technologies, the sub threshold leakage current, I_{SUB}, is much larger than the other leakage current components [5]. This is mainly because of the relatively low V_T in modern CMOS devices. I_{SUB} is calculated by using the equation (1)

$$I_{SUB} = \frac{W}{L}\mu V_{th}^2 C_{STH} e^{\frac{Vgs-Vt+\eta Vds}{nVth}}\left(1-e^{\frac{-Vds}{Vth}}\right)........1$$

where W and L denote the transistor width and length, μ denotes the carrier mobility, $V_{TH} = \frac{kt}{q}$ is the thermal voltage at temperature T, C_{STH} = Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling

and grammar:

C_{DEP}+ Citdenotes the summation of the depletion region capacitance and the interface trap capacitance both per unit area of the MOS gate , and η is the drain-induced barrier lowering (DIBL) coefficient [6]nis the slope shape factor and is calculated by equation-2

$$n = 1 + \frac{Csth}{Cox} \dots \dots \dots \dots \dots \dots \ 2$$

3.3 OXIDE TUNNELING CURRENT (IG)

Reduction of gate oxide thickness results in an increase inthe field across the oxide. The high electric field coupled with low oxide thickness results in tunneling of electrons from substrate to gate and also from gate to substrate through the gate oxide, resulting in the gate oxide tunneling current. Ignoring the effect of finite temperature and image-force-induced barrier lowering, the current density in the FN tunneling is given by equation-3

$$J_{FN} = \frac{q^3 E^2 ox}{16\pi^2 h \emptyset ox} e^{-\frac{\emptyset^{3/2} ox \ 4\sqrt{2m}}{3hqEox}} \ \dots \dots .3$$

Where E_{OX} is the field across the oxide; $\emptyset_{OX}$ is the barrier height for electrons in the conduction band; and m^* is the effective mass of an electron in the conduction band of silicon. The FN current equation represents the tunneling through thetriangular potential barrier and is valid for $V_{OX} > \emptyset_{OX}$, where is the voltage drop across the oxide

3.4 GATE CURRENT DUE TO HOT-CARRIER INJECTION

In a short-channel transistor, due to high electric field near the Si–SiO₂ interface, electrons or holes can gain sufficient energy from the electric field to cross the interface potential barrier and enter into the oxide layer. The injection from Si to SiO₂ is more likely for electrons than holes, as electrons have a lower effective mass than that of holes and barrier height for holes (4.5ev) is more than that for electron

3.5 GATE INDUCED DRAIN LEAKAGE

GIDL is due to high field effect in the drain junction of an MOS transistor. When the gate is biased to form an accumulation layer at the silicon surface, the silicon surface under the gate has almost the same potential as the p-type substrate. Due to presence of accumulated holes GIDL is due to high field effect in the drain junction of an MOS transistor. When the gate is biased to form an accumulation layer at the silicon surface, the silicon surface under the gate has almost same potential as the p-type substrate. Due to presence of accumulated holes.

3.6 CHANNEL PUNCH THROUGH CURRENT USING THE TEMPLATE

In short-channel devices, due to the proximity of the drain

and the source, the depletion regions at the drain-substrate and source-substrate junctions extend into the channel. As the channel length is reduced, if the doping is kept constant, the separation between the depletion region boundaries decreases. An increase in the reverse bias across the junctions (with increase in V_{DS}) also pushes the junctions nearer to each other.

4. LEAKAGE REDUCTIONS AT PROCESS LEVEL

In an SRAM cell, the total power dissipation in dynamic and static components during the active mode. In the standby mode, the power dissipation is due to the standby leakage current. Dynamic power dissipation consists of two components. One is the switching power due to charging and discharging of load capacitance. The other is short circuit power due to the nonzero rise and fall time of input waveforms. The static power of a CMOS circuit is determined by the leakage current through each transistor. The dynamic (switching) power (P_D) and leakage power (P_{LEAK}) are expressed as

$$P_D = \alpha f C V_{DD}^2 \dots \dots \dots \dots \dots \dots \ 4$$
$$P_{LEAK} = I_{LEAK} . V_{DD} \ \dots \dots \dots \dots \ 5$$

Where α is the switching activity; f is the operation frequency; C is the load capacitance. The substrate doping concentration should increase to decrease the depletion width proportionally. The principle of constant field scaling lies in scaling the device voltages and the device dimensions by the same factor, In addition to gate oxide thickness and junction scaling, another technique to improve short-channel characteristics is well engineering. By changing the doping profile in the channel region, the distribution of the electric field and potential contours can be changed. The goal is to optimize the channel profile to minimize the OFF-state leakage while maximizing the linear and saturated drive currents.
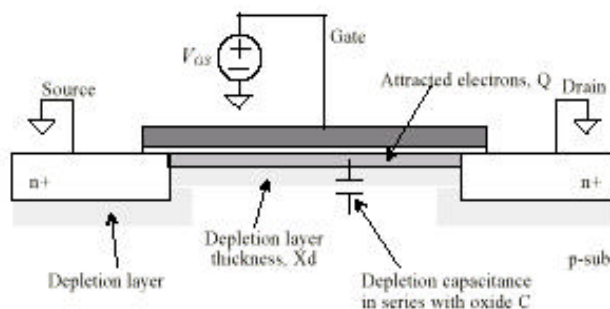

Fig. 5: MOSFETs Graphical Representation

5. LEAKAGE REDUCTIONS AT CIRCUIT LEVEL

There are various techniques for leakage reduction at circuit level. To reduce the leakage currents, the use of higher V_T. Transistor should be influenced through the bulk and DIBL coefficients. As these coefficients is only

technology. For the gate leakage current only the voltages across the gate oxide can be adjusted. There are some methods.

5.1 SELF REVERSE BIASING

The body effect in CMOS transistors, a smaller width of the depletion layer leads to lower VT. The reverse biasing of CMOS transistor increases VT while on forward biasing of the COMS transistor VT decreases. And also in CMOS threshold voltage increases with increased doping of the channel but decreases with applied bias. Therefore the current in the sub threshold region can be partially decreased by reverse biasing.

Equation (6) quantifies the back-gate bias parameter as function of the oxide capacitance and substrate doping level [7].

$$\gamma = \frac{t_{ox} \cdot \sqrt{2 N_{SUB} q \epsilon_{Si}}}{\epsilon_{ox}} \dots \dots \dots (6)$$

Where t_{ox} is gate oxide thickness, N_{SUB} is substrate doping level q is unity electron charge ϵ_{ox} is gate oxide permittivity and ϵ_{Si} is Silicon permittivity.

Fig. 6: MOS Transistor With Reversed Biased

5.2 MULTIPLE VTH DESIGNS

Multiple-threshold CMOS technologies is most popular and easy technique which put high V_T and low V_T transistors in a same chip, very effectively in leakage problem. The high-threshold transistors can partially decreased sub threshold leakage current, while the low-threshold transistors are used for high performance and Stability.

Multiple-threshold voltages can be achieved by the following methods.

1) Multiple Channel Doping: The threshold voltage equals the sum of the flat band voltage, twice the bulk potential and the voltage across the oxide due to the depletion layer charge.

$$V_T = V_{FB} + 2\varphi_F + \frac{\sqrt{2\epsilon_q q N_a (2\varphi_F + V_{SB})}}{C_{OX}} \dots \dots \dots (7)$$

Where the flat band voltage, V_{FB} is given by

$$V_{FB} = \varphi_{FB} - \frac{Q_F}{C_{OX}} - \frac{1}{C_{OX}} \int_0^{tox} \frac{x}{x_{OX}} \rho_{OX}(x) dx \dots \dots (8)$$

With $\quad \varphi_{MS} = \varphi_M - \varphi_S = \varphi_M - (x + \frac{E_g}{2q} + \varphi_F)$

And $\quad \varphi_F = V_t \ln \frac{N_a}{n_i}, p - substrate$

And the similar for pMOS

$$\varphi_F = V_t \ln \frac{N_d}{n_i}, n - substrate$$

The threshold voltage dependence on the doping density is illustrated with for both *n*-type and *p*-type MOSFETs with an aluminum gate metal. The threshold of both types of devices is slightly negative at low doping densities and differs by 4times the absolute value of the bulk potential. The threshold of nMOSFETs increases with doping while the threshold of pMOSFETs decreases with doping in the same way.
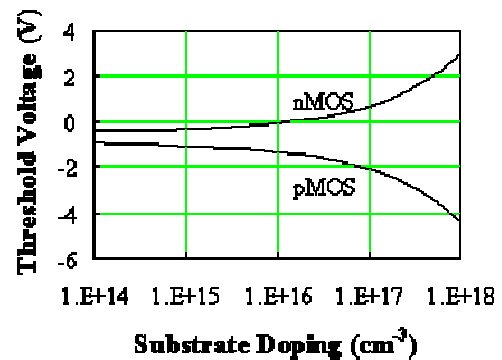

Fig. 7: Threshold voltage of *n*-type (upper curve) and *p*-type (lower curve) MOSFETs versus substrate doping density

A variation of the flat band voltage due to oxide charge will cause a reduction of both threshold voltages if the charge is positive and an increase if the charge is negative.

2) Multiple Oxide CMOS (M_{OX}): Gate oxide thickness can be used to modify the threshold voltage of a transistor Dual can be achieved by depositing two different oxide thicknesses lower oxide thickness, and hence lower threshold voltage, in critical paths maintains the performance. Higher oxide thickness not only reduces the sub threshold leakage, it also reduces: a) gate oxide tunneling, since the oxide tunneling current exponentially decreases with an increase in the oxide thickness [8]; b) dynamic power consumption, since higher oxide thickness reduces the gate capacitance, which is beneficial for reduction of the dynamic power.
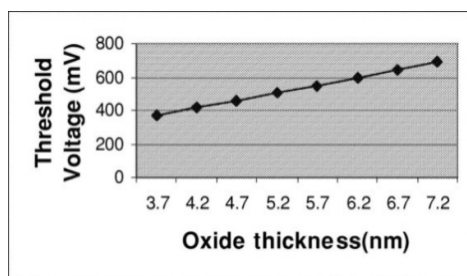
Fig. 8: VTH at Different Oxide Thicknesses

3) Multiple Threshold Voltage: Employing transistors of both high and low V_T values in a single chip along with a sleep control mechanism is the easiest method of reducing static power component. Use of higher V_T transistors outside the critical path reduces static power consumption without compromising operational speed.

4) Multiple Supply Voltages: Applying different supply voltages is another way of reducing the leakage currents because the drain-source voltage of PMOS and NMOS transistors are closely linked to the supply voltage and power dissipation decreases quadratically with the scaling of supply voltage. Although the reduction of power supply voltage significantly reduces the dynamic power dissipation the inevitable design tradeoff is the increase of delay. Since the delay is proportional to $V_{DD}/(V_{DD} - V_T)^2$ it is possible to use high supply voltage in the critical paths of a design to achieve the required performance while the off critical paths of the design use lower supply voltage to achieve low power dissipation.

5) Dual Supply SRAMs: In CMOS SRAMs there are two important performance metrics, fast accessing speed and retention of the stored data. To fulfill this objective using two supplies to allow the SRAM to run in two different modes. A nominal supply voltage is used for powering of the peripheral circuits as decoders, sense-amplifiers and control circuitry and a secondary lower supply is used for idle cells. This effectively lowers the leakage currents of the idle cells. This idea can be utilized in many different ways. In [9] whole banks are put in a drowsy mode and require one or more clock cycles to be accessed. Resultant power saving are predicted in the order of 50% to 75%.

6.   SIMULATION AND RESULT
We simulate various SRAM cell on cadence tool in different-different mode and technique and circuit parameter in 6T SRAM and 7T SRAM and found that 7T is the most prominent low power consumption cell. The fig-9 is shown hybrid 7T-SRAM cell with use various circuit technique and process. The leakage power is reduced by high level using various techniques. The fig-10 showed the leakage current and power with compare of 6T SRAM cell and 7T SRAM cell with also process and circuit level techniques modeling. Its also compare with hybrid model of 7T SRAM cell.

We have clear that V_T is the most appropriate parameter for reducing leakage power and current. By increasing doping and decreasing T_{OX} reduced power consumption.

The Equation-7 show that the threshold voltage dependent on various parameters on SRAM cell. Here we can see that the threshold can changed by Ø, γ and η. They are easily change by well engineering design.

$$V_T = V_{TO} + \gamma(\sqrt{(-2\emptyset f)} - V_{SB} - \sqrt{2\emptyset}\dotso 7$$
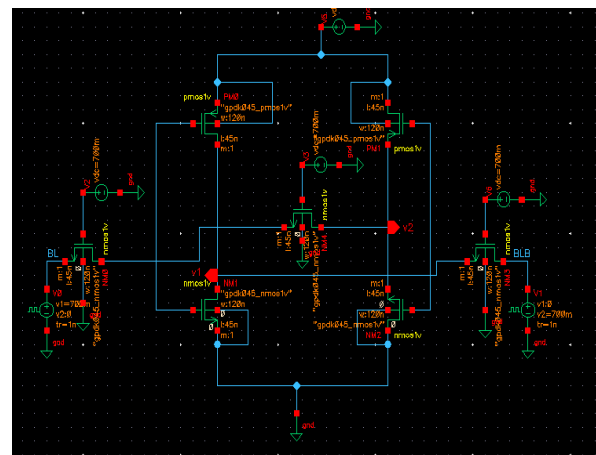

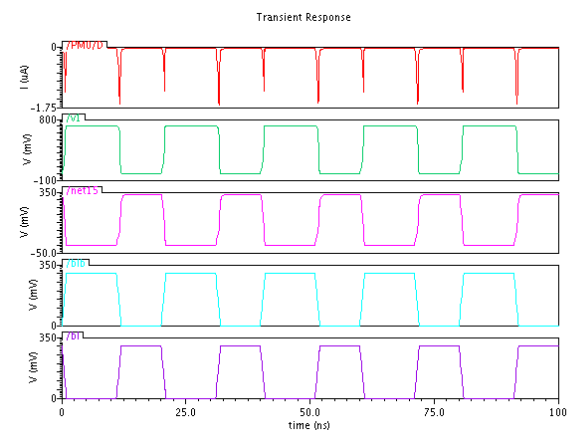**Fig. 9: Used High Level circuit and process technique Simulated 7T SRAM**
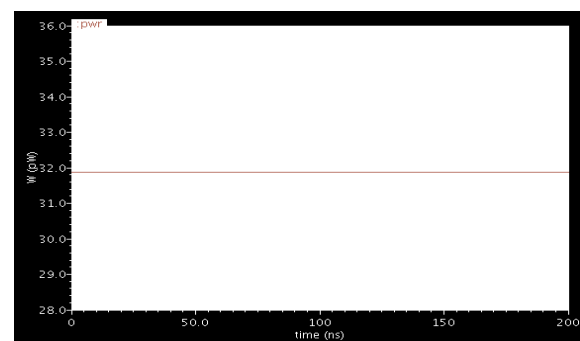

Fig. 10: Leakage current for 7T SRAM

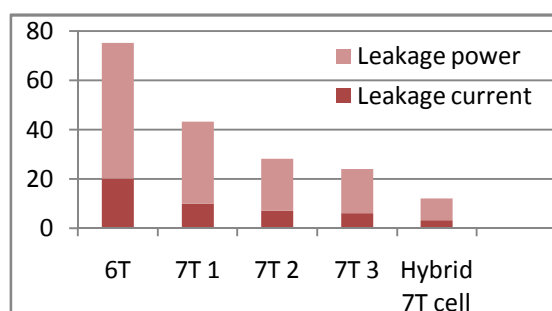**Fig. 11: Leakage power for 7T SRAM**

Fig. 12: Various result of Current and Power

7. ACKNOWLEDGEMENT

8. REFERENCES

[1] AminulIslam and Mohd. Hassan "variability analysis of 6t and 7t sramcell in sub-45nm technology" IIUM engineering journal, vol. 12, no. 1, 2011

[2] Shilpi Birla, R.K.Singh, Member IACSIT, and Manisha Pattnaik, Static Noise Margin Analysis of Various SRAM Topologies, IACSIT International Journal of Engineering and Technology, Vol.3, No.3, June 2011

[3] Debasis Mukherjee, Hemanta Kr. Mondal and B.V.R. Reddy," Static Noise Margin Analysis of SRAM Cell for High Speed Application" IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010.

[4] Benton H. Calhoun, Member, IEEE, and Anantha P. Chandrakasan, Fellow," Static Noise Margin Variation for Sub-threshold SRAM in65-nm CMOS" IEEE, IEEE Journal Of Solid-State Circuits, Vol. 41, No. 7, July 2006.

[5] Sanjeev K. Jain ,Pankaj Agarwal "A Low Leakage and SNM Free SRAM Cell Design in Deep Sub micron CMOS Technology" Proceedings of the 19th International Conference on VLSI Design (VLSID'06)

[6] Andrei Pavlov &ManojSachdev, "CMOS SRAM Circuit Design and Parametric Test in Nano-Scaled Technologies". Intel Corporation, University of Waterloo, 2008 Springer Science and Business Media B.V., pp:1 2002.

[7] Rafik S. Guindi , Farid N. Najm, Design Techniques for Gate-Leakage Reduction in CMOS Circuits, Proceedings of the Fourth International Symposium on Quality Electronic Design, p.61, March 24-26, 2003.

[8] B. Alorda, G. Torrens, S. Bota and J. Segura Univ. de les IllesBalears," Static-Noise Margin Analysis during Read Operation of 6T SRAM Cells" Dept. Fisica, Cra. Valldemossa, km. 7.5, 07071 Palma de Mallorca, Spain.

❑❑❑

IMAGE ENRICHMENT USING ANFIS TECHNIQUES

[1]Sunil Kumar Yadav* and [2]Shahina Anjum
[1,2]Assistant Professor, Department of Computer Science and Engineering
Accurate Institute of Management & Technology, Greater Noida, U.P., India
*Corresponding Author : sunil.3232@gmail.com

ABSTRACT

Image enrichment plays an essential role in various applications. The principle objective of Image enrichment is to process an image so that result is more suitable than the original image for specific application. There are many techniques to remove the noise from the image and produce the clear visual of the image. Moreover, there are several filters and image smoothing techniques available in the literature. All these available techniques have certain limitations. Recently, neural networks have been found to be a very efficient tool for image enrichment. Here we are borrowing the concepts of fuzzy domain image enrichment method along with the learning abilities of neural networks to propose a hybrid neuro-fuzzy image enrichment methodology. ANFIS is a fuzzy inference system (FIS) implemented in the framework of an adaptive fuzzy neural network. It combines the explicit knowledge representation of an FIS with the learning power of artificial neural networks. The objective of ANFIS is to integrate the best features of fuzzy systems and neural networks. This paper proposes an adaptive neuro-fuzzy inference system (ANFIS) based method for the enrichment of images corrupted by salt and pepper noise. After the noise detection, the filtering is performed. In this paper, we will implement our work on MATLAB 7.9. The proposed technique is able to improve the contrast of the image using the histogram processing technique.

Keywords : Image Enrichment, Noise and Filters, ANFIS, Contrast Enrichment, Histogram Processing.

1. INTRODUCTION

The main goal of image enhancement is to process an image so that the result is more appropriate than the original image for a particular application. It sharpens image features such as edges, boundaries, or contrast to make a graphic display more helpful for display and analysis. The enhancement doesn't boost the basic information content of the data, but it increases the dynamic range of the chosen characteristics so that they can be found easily. The main difficulty in image enhancement is quantifying the criterion for enhancement and, therefore, a large number of image enhancement techniques are empirical and need enhancement process to obtain satisfactory results. Image enrichment methods can be based on either spatial or frequency domain techniques. [1].

In this paper, we focus on the spatial domain techniques which are performed to the image plane itself and they are based on direct manipulation of pixels in an image. The operation can be formulated as:

$$g(x, y) = T[f(x, y)] \qquad (1)$$

Where g is the output, f is the input image and T is an operation on f defined over some neighborhood of (x, y). According to the operations on the image pixels, it can be divided into two parts: Point operations and spatial operations [1].
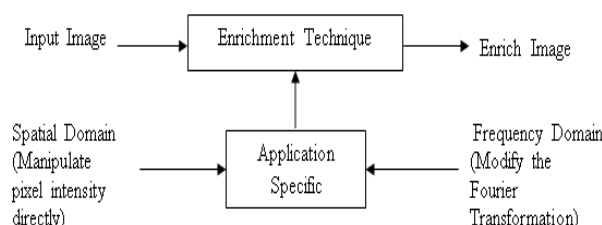

Fig. 1: Basic Model of Image Enrichment [1]

This paper aim to provide basic concepts of image enhancement, along with algorithms which is commonly used for image enhancement. The paper focuses on spatial domain techniques for image enrichment, with particular reference to histogram processing.

1.1 NOISE AND FILTERS

I. IMAGE NOISE

Image noise is the arbitrary variation of brightness or color information in images produced by the sensor and circuitry of a digital camera or scanner. Image noise can also begin in film grain and in the unavoidable shot noise of a perfect photon detector. Image noise is generally

called as an undesirable by-product of image capture. Although these undesirable fluctuations became known as "noise" by analogy with undesirable sound they are inaudible and such as dithering[3].

The various types of noise are:
- Amplifier noise (Gaussian noise)
- Salt-and-pepper noise
- Shot noise (Poisson noise)
- Speckle noise

II. FILTERS

Filters are used for removing the noise, the process is called as image-denoising. Image de-noising is a vital image processing task, i.e. as a process itself as well as a component in other processes. There are many ways to de-noise an image or a set of data and methods exists. The important property of a good image denoising model is that it should completely remove noise as far as possible as well as preserve edges. Traditionally, there are two types of models, i.e. linear model and non-linear model. Generally, linear models are used. The benefits of linear noise removing models is the speed and the limitations of the linear models is, the models are not able to preserve edges of the images in a efficient manner i.e the edges, which are recognized as discontinuities in the image, are smeared out. On the other hand, Non-linear models can handle edges in a much better way than linear models. [3] Some of the noise filter models are:
- Mean Filter Module
- Median Filter Module
- Wiener Filter Module

In this paper, the work is done to de-noising the salt and pepper noise where, an image containing salt-and-pepper noise will have dark pixels in bright regions and bright pixels in dark regions. This type of noise can be caused by dead pixels, analog-to-digital converter errors, bit errors in transmission, etc. This can be eliminated in large part by using dark frame subtraction and by interpolating around dark/bright pixels. After detecting the noise, we proposed the use of filter on the image containing the salt and pepper noise for the image de-noising process.

1.2 ANFIS (ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM)

Neural networks are used to design membership functions of fuzzy systems that are employed as decision-making systems for controlling equipment. However, fuzzy logic can encode expert knowledge directly by using rules with linguistic labels, it usually takes a lot of time to design and tune the membership functions which quantitatively define these linguistic labels. Neural network learning methods can mechanize this process and considerably reduce development time and cost while increasing the performance. The idea of using neural networks to design membership functions was proposed by Takagi and Hayashi.

I. ANFIS Structure

For simplicity, we assume a network with two inputs, x and y, and one output, f. The ANFIS is a fuzzy Sugeno model. To present the ANFIS architecture,[5] two fuzzy if–then rules based on a first-order Sugeno model are considered:

Rule 1: if x is A_1 and y is B_1, then $f_1 = p_1x + q_1y + r_1$
Rule 2: if x is A_2 and y is B_2, then $f_2 = p_2x + q_2y + r_2$

Where, x and y are the inputs, A_i and B_i are the fuzzy sets, f_i, i= 1, 2 are the outputs of the fuzzy system and p_i, q_i and r_i are the design parameters that are determined during the training process. The ANFIS architecture to implement these two rules is shown in Fig 1.2 in which a circle indicates a fixed node whereas a square indicates an adaptive node. As illustrated in the figure, the ANFIS architecture consists of five layers.
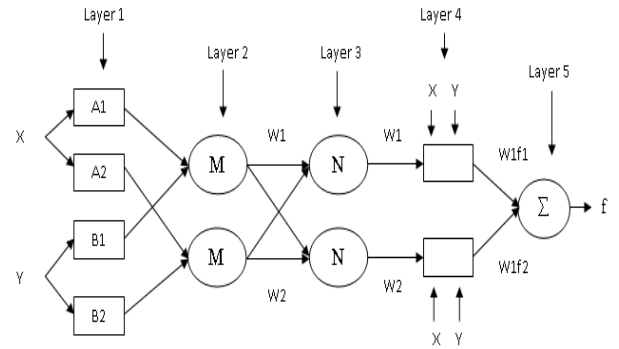

Fig. 2: Adaptive neuro-fuzzy inference system structure

Layer 1: Every node in layer 1 is an adaptation node. The outputs of layer 1 are the fuzzy membership grades of the inputs, which are given by:

$$O_i^1 = \mu Ai\,(x) \quad , \qquad i=1,2 \qquad (2)$$
$$O_i^1 = \mu Bi - 2(y) \quad , \quad i=3,4 \qquad (3)$$

where, x and y are the inputs to node i, A is a linguistic label and $\mu Ai\,(x)$ and $\mu Bi{-}2\,(y)$, can adopt any fuzzy membership function. Usually, $\mu Ai\,(x)$ can be selected such as:

$$\mu_{Ai}\,(x) = \frac{1}{1+\{\left(\frac{x-c_i}{a_i}\right)^2\}^{b_i}} \qquad (4)$$

Where, a_i, b_i and c_i are the parameters of the membership bell-shape function.

Layer 2: The nodes of this layer are labeled M, indicating that they perform as a simple multiplier. The outputs of this layer can be represented as:

$$O_i^2 = Wi = \mu Ai\,(x)\,\mu Bi\,(y) \,, i=1, 2 \quad (5)$$

Layer 3: It contains fixed nodes that calculate the ratio of the firing strengths of the rules as follows:

$$O_i^3 = W_i = \frac{W_i}{(W_1 + W_2)} \quad, \quad i=1, 2 \ (6)$$

Layer 4: In this layer, the nodes are adaptive nodes. The outputs of this layer are computed by the formula given below:

$$O_i^4 = \overline{W_i} f_i = \overline{W_i}(p_i x + q_i + r_i), i=1, 2 \ (7)$$

where, $\overline{W_i}$ is a normalized firing strength from layer 3.

Layer 5: The node performs the summation of all incoming signals. Hence, the overall output of the model is given by:

$$O_1^5 = \sum_{i=1}^{2} \overline{w}_i f_i = \frac{\sum_{i=1}^{2} w_i f_i}{\sum_{i=1}^{2} w_i} \qquad (8)$$

It can be seen that there are two adaptive layers in this ANFIS architecture, namely the first layer and the fourth layer. In the first layer, there are three modifiable parameters $\{a_i, b_i, c_i\}$, which are related to the input membership functions. These parameters are generally termed as premise parameters. In the fourth layer, there are also three modifiable parameters $\{p_i, q_i, r_i\}$ pertaining to the first-order polynomial. These parameters are the so-called consequent parameters.

2. RELATED WORK

M. Emin Yükse (2006) [6] presented a new operator for restoring the images which are corrupted by the impulse noise. The proposed operator is an appropriate combination of a median filter, an edge detector, and a neuro-fuzzy network and performed as a hybrid filter. The internal parameters of the neuro-fuzzy network are adaptively optimized by training. The training is easily accomplished by using simple artificial images that can be generated in a computer. The most distinctive feature of the proposed operator over most other operators is that it offers excellent line, edge, detail, and texture preservation performance while, at the same time, effectively removing noise from the input image. Extensive simulation experiments show that the proposed operator may be used for efficient restoration of digital images corrupted by impulse noise without distorting the useful information in the image.

Kundra, et al. (2009) [7] studied that the imaging and transmission is the process where to avoid the interference of various kinds of noise is very hard. So to overcome this problem, the preprocessing steps like image enrichment are widely used in the presence of noise. To enhance the main structures in the input image and to remove the impulsive noise, the image enrichment techniques are used. The image smoothing and image sharpening are the two important methods used in this technique. In this paper the two step algorithm is used by removing the impulse noise in a first step and in the second step, the contrast of the image is improving. To achieve this goal fuzzy-logic-control based approach is used. The filter is tested on the colored images. The output generated by this method is an enhanced image with no noise. The contrast is also improved. Thus, the above mentioned method proves the best approach for enriching the corrupted images.

Claudia Nieuwenhuis et al. [8] In this paper the concept of adaptive filters with neural networks was combined in order to be able to include high level knowledge about the contents of the image in the filtering process. The survey said that the Adaptive image Enrichment algorithms often utilize low level knowledge like gradient information to guide filtering parameters. The independency of these filters on any specific knowledge and their applicability on a broad spectrum of images is the main advantage for use. However, for many problems this low level information is not sufficient to achieve good results. For example, in medical imaging it is often very important that some features are preserved while others are suppressed. Usually these features cannot be distinguished by low level information. Therefore the proposed method is to incorporate high level knowledge in the filtering process in order to adjust the parameters of any given filter thus creating a guided filter.

George C. Giakos et al. [9] In this paper, basic neural network algorithms was applied to the imaging process as well as their applications in different areas of technology, are presented, discussed, and analyzed. Novel ideas towards the optimization of the design parameters of digital imaging sensors utilizing neural networks are presented. Several problems associated with low-contrast images, blurred images, noisy images, image conversion to digital form, transmission, handling, manipulation, and storage of large-volume images, led to the development of efficient image processing and recognition algorithms. Digital imaging or computer vision involves image processing and pattern recognition techniques Image processing techniques deal with image Enrichment, manipulation, and analysis of images.

3. PROPOSED SYSTEM ARCHITECTURE

In this paper, we basically perform the test on the jpeg format images which will convert into gray scale format before processing. The paper focuses on spatial domain techniques for image enrichment, with particular reference to histogram processing. This all is done on the MATLAB7.9 platform.

The proposed system is divided into two stages.

3.1 NOISE REMOVAL

In this step we perform to remove the noise by applying the filters. In our work the focus is on the salt-and pepper noise which we remove by the use of filters.

3.2   IMAGE ENRICHMENT

This paper proposes an adaptive neuro-fuzzy inference system (ANFIS) based method for the enrichment of images corrupted by salt and pepper noise. After the noise removal Image Enrichment is done to improve the visual quality of an output image. The objective of ANFIS is to integrate the best features of fuzzy systems and neural networks. The proposed technique is able to improve the contrast of the image using the histogram processing technique.

A. PROPOSED ALGORITHM

Step 1: Read the image.
Step 2: If the image is colored then convert the image into a gray level image.
           Add noise to the image.
Step 3: The size of gray image is m x n where I={I(i,j) e(0,1,…255)}
Step 4: Select an image window matrix p x q.
Step 5: Fuzzification of the image pxq matrix is done
Step 6: Fuzzy rules are created for enriching the image
Step 7: Apply filters on enriched image
Step 8: Defuzzify the output. Let neural network learn the parameters to adjust the weights.
Step 9: Compare the enriched image with other image.

B. CONTRAST ENRICHMENT WITH HISTOGRAM PROCESSING

I. CONTRAST ENRICHMENT

Contrast enrichments improve the perceptibility of objects in the scene by enriching the brightness difference between objects and their backgrounds. Contrast enrichments are typically performed as a contrast stretch followed by a tonal enrichment, although these could both be performed in one step. A contrast stretch improves the brightness differences uniformly across the dynamic range of the image, whereas tonal enrichments improve the brightness differences in the shadow (dark), midtone (grays), or highlight (bright) regions at the expense of the brightness differences in the other regions. [10]

II. HISTOGRAM PROCESSING

The histogram of a digital image with intensity levels in the range [0, L-1] is a discrete function

$$h\left(r_k\right) = n_k \qquad (9)$$

where r_k = k[th] intensity value and
         n_k = Number of pixels in the image with intensity r_k

Histograms are frequently normalized by the total number of pixels in the image. Assuming a M x N image, a normalized histogram.

$$p(r_k) = \frac{n_k}{MN} \text{, k} = 0,1,…L\text{-}1 \qquad (10)$$

is related to probability of occurrence of r_k in the image.[2]

The shape of the histogram of an image gives us useful information about the possibility for contrast enrichment. A histogram of a narrow shape indicates little dynamic range and thus corresponds to an image having low contrast. [1]

4.   EXPERIMENTAL RESULTS

For experimenting the proposed technique, several gray scale images affected by the noise is considered and the purpose is to enrich the contrast by showing the effect through histogram presentation.
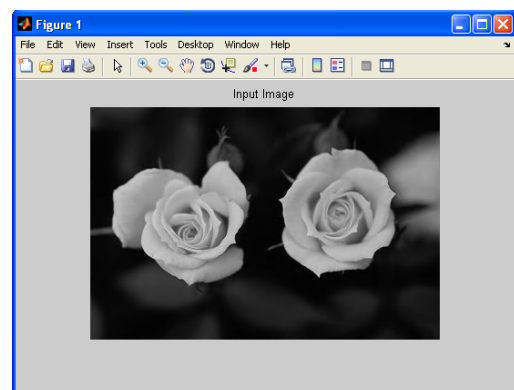
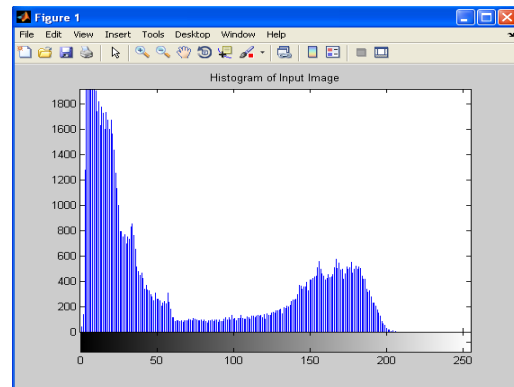## 4.1   SNAPSHOTS OF OUTPUTS FOR 1ST INPUT IMAGE

Fig. 5.1: Input Image (A)
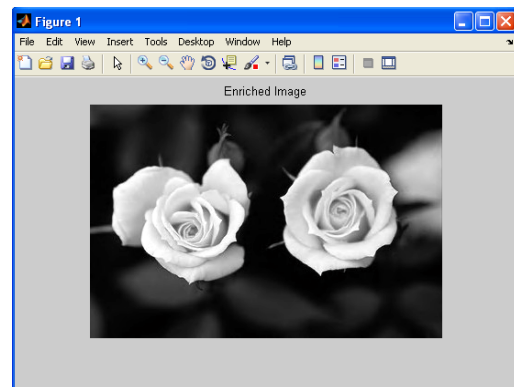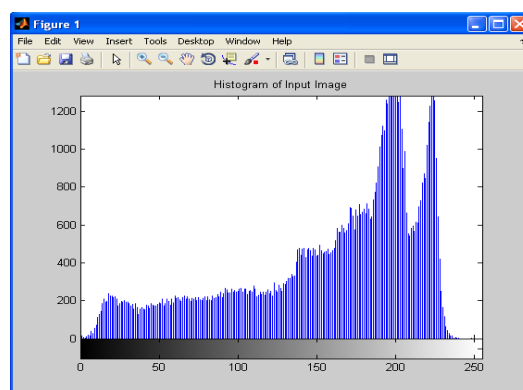

Fig. 5.2: Histogram of Input Image (A)
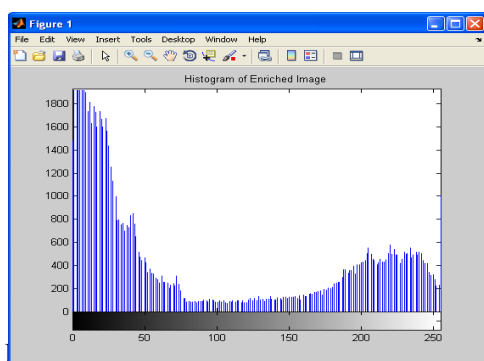

Fig. 5.3: Output Image or Enriched Image (A)
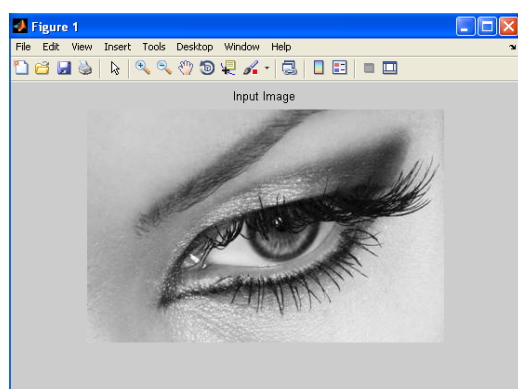
4.2 SNAPSHOTS OF OUTPUTS FOR 2ND INPUT IMAGE

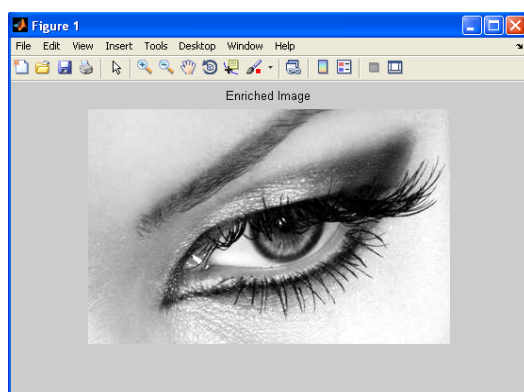
Fig. 5.5: Input Image (B)
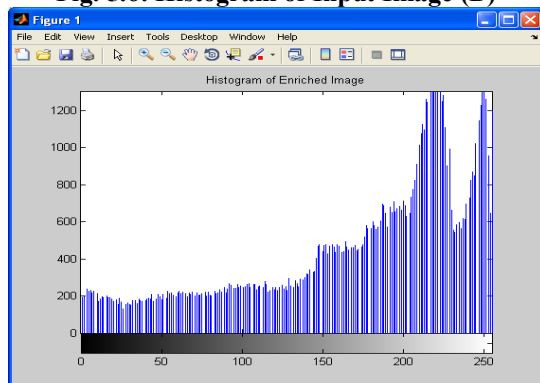

Fig. 5.6: Histogram of Input Image (B)

Fig. 5.7: Output Image or Enriched Image (B)

5. CONCLUSION

Image enrichment is the improvement of digital image quality, where technique can be processed without having knowledge about the source of degradation. A two- stage architecture is proposed in this paper where in the first stage a noise is removed by using the filter and then Adaptive Neuro-Fuzzy Inference System was applied on the first stage output images for the enrichment in the final output images. The above are the test images as well as the output generated by applying the method. The outputs were clearly shown in the output window of the MATLAB platform, the output generated is an enhanced image with no noise. The contrast is also improved and by observing the histogram of input and output images, it is clear that the output image is high-contrast image as it covers all the possible values in the gray scale used in histogram whereas a histogram of a narrow shape indicates little dynamic range and thus corresponds to an image having low contrast which is of the input image in our snapshots. This paper concludes that use of the ANFIS (adaptive neuro-fuzzy inference system) proves as the better approach for enhancing the corrupted images.

6. REFERENCES

[1] Source: http://www.eie.polyu.edu.hk/~enyhchan/imagee.pdf

[2] Raman Maini and Himanshu Aggarwal, "A Comprehensive Review of Image Enrichment Techniques"Journal of Computing, Volume2. issue3, March-2010

[3] Pawan Patidar, Manoj Gupta Sumit Srivastava,Ashok Kumar Nagawat," Image De-noising by Various Filters for Different Noise", International Journal of Computer Applications (0975 – 8887) Volume 9– No.4, November 2010

[4] Mahesh T R, Prabhanjan S, M Vinayababu, "Noise Reduction by using Fuzzy Image Filtering ",Journal of Theoretical and Applied Information Technology, Islamabad PAKISTAN,31st May 2010. Vol.15. No.2.

[5] Monireh Sheikh Hosseini ,Maryam Zekri ,"Review of Medical Image Classification using the Adaptive Neuro-Fuzzy Inference System", J Med Signals Sens. 2012 Jan-Apr; 2(1): 49–60.

[6] M. Emin Yüksel, " A Hybrid Neuro-Fuzzy Filter for Edge Preserving Restoration of Images Corrupted by Impulse Noise," IEEE Transactions on Image Processing, Vol. 15, No. 4, April 2006

[7] Harish Kundra, Er. Aashima, Er. Monika Verma "Image Enrichment Based On Fuzzy Logic" IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.10, October 2009

[8] Claudia Nieuwenhuis, Michelle Yan, "Knowledge Based Image Enrichment Using Neural Networks",Technical University of Ilmenau ,Department of Neuroinformatics, Siemens Corporate Research

[9] George C. Giakos , Kiran NATARAJ, Ninad PATNEKAR, "Neural Networks for Image Analysis and Processing in Measurements, Instrumentation and Related Industrial Applications", The University of Akron OH 44325-3904, USA

[10] Source: http://spie.org/samples/TT92.pdf

❑❑❑

International VSRD

J O U R N A L S

A Research Division of Visual Soft (India) Pvt. Ltd.