S.no	Title of the paper	Name of the author	Title of the journal	Year of publication	Citation Index	Institutional affiliation as mentioned in the publication	Number of citations excluding self citations
	2014-2015						
1	Ordering and Suggesting Popular Itemsets in Pharmacy using Modified Apriori Algorithm	A Naga Sri, B Sujatha, CH Bhavani, B Ravinder Reddy, B Satish Gupta	IJIREEICE	Feb 2014	ISSN 2321-2004	-	-
2	Routing Framework for Delay Tolerant Networks using Bayesia Lab	A Naga Sri, B Sujatha, CH Bhavani	IOSRJCE	Feb 2014	e-ISSN:2278-0661		
3	Efficient Detection of Internet worms using Data Mining Techniques	B Sujatha, G Rajitha Devi	IOSRJCE	Mar-Apr 2014	e-ISSN: 2278-0661		
4	Efficient Ranking and Suggesting ℶopular items in Mobile Stores using Fp Tree Approach	B Sujatha, Shaista Nousheen, Tasneem Rahath, Nikhath Fatima	IJCER	May 2014	e-ISSN: 2250-3005		
5	Improving Maximal Frequent Item set mining for Sparse Dataset	B Sujatha, V Ramesh Babu	IJSER	April 2014	ISSN 2229-5518		
6	Enterprise Resource Planning – Analysis of Business Intelligence and Emergence of Mining Objects	Ramesh Babu Varugu, Asst Prof	IJETTCS	May-June2014	ISSN: 2278-6856	Annamacharya Institute of Science and Technology	
7	Tree Based Graph Mining – Analysis of Interaction Pattern Discovery in Business	Ramesh Babu Varugu, C Shanker,	IJEEE	August 2014	ISSN: 2348 - 4748	Annamacharya Institute of Science and Technology, Sri Indu College of Engineering & Technology	
	ꞏsign & Development of an Effective Video Streaming Framework using Cloud Computing Technology	K Narasimhulu, K V S Sudhakar, N Yadagiri, Prof. Dr. G. Manoj 8Someswar	IJSRCSAMS	Nov 2014	ISSN: 2319 - 1953		
	Analysis of Cloud Data Mining & Emergence of Self Destructing Technique to Archive Data	Ramesh Babu varugu	IJETCR	Dec 2014	ISSN: 2348 - 2117		
	2015-2016						
	Cloud data center management with Quality of service using a Novel Stochastic Mode	B Satish gupta, Shafiulilah Shaik	IJSETR	January 2015	ISSN: 2319 - 8885	-	-

PRINCIPAL
Annamacharya Institute of
Technology & Sciences
Piglipur, Batasingaram (V),
Hayathnagar (M) R.R. Dt., 501 512

	2016-2017						
1	Enabling Fine-Grained Multi-keyword Search Supporting Classified Sub-Dictionaries over Encrypted Cloud Data	G Sandhya Rani; Ramesh Babu Varugu; V Vedasahithi	IJITECH	September 2016	ISSN: 2321-8665	Annamacharya Institute of Technology and Sciences	
2	Improved Privacy preserving P2P Multimedia distribution based on recombined finger prints	Jahnavi Parvathaneni, A Naga Sri, V Ramesh Babu	IJITECH	September 2016	ISSN: 2321 – 8665	Annamacharya Institute of Technology and Sciences	
3	A Secure Anti-Collusion Data Sharing Scheme for Dynamic Goups in the Cloud	Anaganti Sudha,V.Ramesh Babu, A.Nagasri	IJITECH	September 2016	ISSN: 2321 – 8665	Annamacharya Institute of Technology and Sciences	
4	CLOUD ARMOR: Supporting Reputation-Based Trust Management for Cloud Services	Dasari Swapna Varugu Ramesh Babu B.Ravinder Reddy	IJITECH	September 2016	ISSN: 2321 – 8665	Annamacharya Institute of Technology and Sciences	
5	Energy-Aware Load Balancing and Application Scaling for the Cloud Ecosystem	A.Sindhuja Varugu Ramesh babu	IJITECH	September 2016	ISSN: 2321 – 8665	Annamacharya Institute of Technology and Sciences	
	2017-2018						
1	Supporting and Classification of Issues in large Descriptive Datasheets	V.Ramakrishna M.Vineela	IJITR	September 2017	ISSN-7386-7388	Annamacharya Institute of Technology and Sciences	
2	A Note on one secure anti collision data sharing scheme for Dynamic groups in the cloud	A.Vanitha V.Ramakrishna	IJR	November 2017	ISSN-2348-6848	Annamacharya Institute of Technology and Sciences	
	Circuit Cipher text Policy attribute based hybrid 3ecnryption with verifiable delegation in cloud computing	P.Navaneetha B.Satish Gupta	IJR	November 2017	ISSN-2348-6848	Annamacharya Institute of Technology and Sciences	
	Nearest Keyword Set Search In Multi Dimensional Datasets	Shaik Nurjahan K.Sandhya Rani	IJR	November 2017	ISSN-2348-6848	Annamacharya Institute of Technology and Sciences	
	A Survey paper on data lineage in malicious environments	Mehabubunnisa K.Nagalatha	IJR	November 2017	ISSN-2348-6848	Annamacharya Institute of Technology and Sciences	
	Spoofer location using passive Ip Trace back	Palde Sudha Jyothi Arava Nagasri	IJR	November 2017	ISSN-2348-6848	Annamacharya Institute of Technology and Sciences	
	Detection of cyber bulling based on automatic code of marginalized denosing improved semantic	Keesara Chamanthi V.Ramesh Babu	IJR	November 2017	ISSN-2348-6848	Annamacharya Institute of Technology and Sciences	

PRINCIPAL
Annamacharya Institute of
Technology & Sciences
Piglipur, Batasingaram (P),
Hayathnagar (M) R.R. Dt., HYD-501 512

8	Exploiting online social for compromised Account Detection	A.Sai Akshitha Devi V.Ramesh Babu	IJR	November 2017	ISSN-2348-6848	Annamacharya Institute of Technology and Sciences	
9	Explore the Malicious Facebook Applications	S.Komali	IJARSE	July 2017	ISSN-2319-8364	Dhruva Institute of Technology & Sciences	
	2018-2019						
1	A Scalable Graph Based Ranking Model for content – based Image Retirval	P.Shiva Prasad, K.Narsimhulu	IJR	May 2018	ISSN-2348-6848	Annamacharya Institute of Technology and Sciences	
2	Survey on Particle Swarm Optimization Techniques in Network-on-Chip	K.Satish Kumar k.paramasivam	IJRTE	November - 2018	ISSN-2277-3878	Annamacharya Institute of Technology and Sciences	
3	Dectecting Financial Fraud by Analyzing Human Behaviour using Data Mining Techniques	M.Jyothi	IJRECS	Jan -2019	ISSN-2321-5784	Annamacharya Institute of Technology and Sciences	

PRINCIPAL
Annamacharya Institute of
Technology & Sciences
Piglipur, Batasingaram (V),
Hayathnagar (M) R.R. Dt., HYD-501 512

International Journal of Research
Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 15
May 2018

Emr: A Scalable Graph-Based Ranking Model for Content-Based Image Retrieval

P Shiva Prasad & K Narsimhulu

*Pg Scholar,** Asst. Professor
Department Of Cse, Annamacharya Institute Of Technology And Sciences Blatasingaram, Hayat Nagar, R.R.Dist, Hyderabad-501512

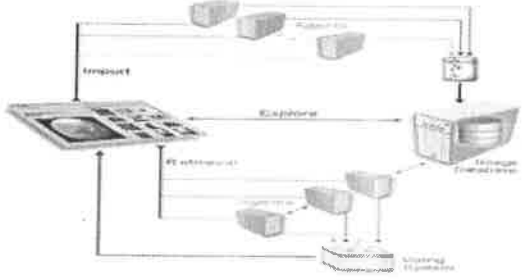
ABSTRACT:

Diagram based positioning models have been generally connected in data recovery zone. In this paper, we concentrate on an outstanding chart based model - the Ranking on Data Manifold model, or Manifold Ranking (MR). Especially, it has been effectively connected to content-based picture recovery, in view of its remarkable capacity to find hidden geometrical structure of the given picture database. Nonetheless, complex positioning is computationally extremely costly, which fundamentally restrains its appropriateness to vast databases particularly for the cases that the inquiries are out of the database (new examples). We propose a novel adaptable chart based positioning model called Efficient Manifold Ranking (EMR), attempting to address the weaknesses of MR from two principle points of view: versatile diagram development and proficient positioning calculation. In particular, we construct a grapple diagram on the database rather than a conventional k-closest neighbor chart, and plan another type of contiguousness network used to accelerate the positioning. A surmised technique is embraced for proficient out-of-test recovery. Test comes about on some huge scale picture databases show that EMR is a promising technique for genuine recovery applications.

INTRODUCTION

Chart BASED positioning models have been profoundly contemplated and generally connected in data recovery range. In this paper, we concentrate on the issue of applying a novel and effective diagram based model for content based picture recovery (CBIR), particularly for out- of-test recovery on expansive scale databases. Conventional picture recovery frameworks depend on watchword seek, for example, Google and Yahoo picture look. In these frameworks, a client catchphrase (question) is coordinated with the setting around a picture including the title, manual explanation, web record, and so on. These frameworks don't use data from pictures. However these frameworks endure numerous issues, for example, lack of the content data and irregularity of the significance of the content and picture. Content-based picture recovery is an extensive decision to defeat these challenges. CBIR has drawn an awesome consideration in the previous two decades. Not the same as conventional catchphrase look frameworks, CBIR frameworks use the low-level highlights, including worldwide highlights (e.g., shading minute, edge histogram, LBP) and neighborhood highlights (e.g., SIFT), naturally removed from pictures. An awesome measure of looks into have been performed for planning more educational low-level highlights to speak to pictures, or better measurements (e.g., DPF) to gauge the perceptual likeness, however their execution is limited by many conditions and is delicate to the information. Importance criticism is a helpful instrument for intuitive CBIR. Client's abnormal state discernment is caught by progressively refreshed weights in light of the client's input. Most conventional strategies concentrate on the information includes excessively yet they disregard the basic structure data, which is of incredible significance for semantic revelation, particularly when the mark data is obscure. Numerous databases have fundamental bunch or complex structure. Under such conditions, the presumption of mark consistency is sensible. It implies that those adjacent information focuses, or directs have a place toward a similar bunch or complex, are probably going to have the same semantic mark. This marvel is critical to investigate the semantic importance when the name data is obscure. As we would see it, a great CBIR framework ought to consider pictures' low-level includes and also the natural structure of the picture database. Complex Ranking (MR), a celebrated chart based positioning model, positions information tests concerning the inherent geometrical structure altogether uncovered by countless. It is precisely in accordance with our thought. MR has been generally connected in numerous applications, and appeared to have astounding execution and practicality on an assortment of information sorts, for example, the content, picture and video. By considering the hidden structure, complex positioning doles out every datum test a relative positioning score, rather than a flat out match shrewd closeness as conventional ways. The score is dealt with as a closeness metric characterized on the complex, which is more important to catching the semantic pertinence degree. He et al.firstly connected MR to CBIR, and altogether enhanced picture recovery execution contrasted and cutting edge calculations. Notwithstanding, complex positioning has its own particular disadvantages to deal with huge scale databases –it has costly computational cost, both in chart development and positioning calculation stages. Especially, it is obscure how to deal with an out-of-test question (another example) proficiently under the current structure. It is unsuitable to re figure the model for another question. That implies, unique complex positioning is deficient for a true CBIR framework, in which the client gave question is dependably an out-of-test. In this paper, we expand the first complex positioning and propose a novel structure named Efficient Manifold Ranking (EMR). We endeavor to address the weaknesses of complex positioning from two viewpoints: the first is versatile diagram development; and the second is proficient calculation, particularly for out-of-test recovery. In particular, we fabricate a grapple diagram on the database rather than the customary k-closest neighbor chart, and plan another type of contiguousness lattice used to accelerate the positioning calculation. The model has two separate stages: a disconnected stage for building (or taking in) the positioning model and an online stage for taking care of another question. With EMR, we can deal with a database with 1 million pictures and do the online recovery in a brief timeframe. To the best of our insight, no past complex positioning based calculation has come up short on test recovery on a database in this scale. A preparatory adaptation of this work beforehand showed up. In this paper, the new commitmentsderation are as to the out-of-test recovery (online stage) and propose an effective rough technique to register positioning scores for another question. Subsequently, we can come up short on test recovery on an expansive scale databasestreamlined theinEMRacode1briefandre-run span. every one of the tests. Three new databases including two huge scale databases with around 1 millions examples are included for testing the effectiveness of the proposed demonstrate. We offer more itemized investigation for trial result. • We formall nearby weight estimation issue for building the grapple diagram and two unique strategies are contrasted with figure out which technique is better. Whatever is left of this paper is sorted out as takes after. we quickly talk about some related work, we survey the calculation of MR and make an examination. The proposed approach EMR. We display the investigation comes about on numerous genuine picture databases.

SYSTEM ARCHITECTURE:



IMPLEMENTATION

- **Admin**

In this module, the Admin needs to login by utilizing substantial client name and secret word. After login effective he can do a few operations, for example, transfer pictures, see transferred pictures, see all informational collections of pictures, rundown of all looking history, see all picture positioning and view all clients, seek pictures and logout.

Transfer Images

In this module, the administrator can transfer n number of pictures. Administrator need to transfer new picture then he has enter a few fields like picture name, picture shading, picture portrayal, picture sort, living spot, peruse the picture document and transfer. In the wake of transferring effectively he will get a reaction from the server. At first new transferred picture rank is zero. In the wake of review that picture rank will re-rank.

View informational collection of Images

In this module, the Admin can see the every one of sort's pictures accessible in server. In the event that administrator needs to see all sort of pictures, at that point tap on informational collection pictures catch, it will offer reaction to client with catchphrases, for example, human, flying creatures, creatures, Insects, organic products, trees and non living articles.

- Hunt History

This is controlled by administrator; the administrator can see the hunt history points of interest. In the event that he taps on seek history catch, it will demonstrate the rundown of sought client points of interest with their labels, for example, client name, client hunt down picture name, time and date.

Rank of pictures

In this module, the administrator can see the rundown of positioning pictures. In the event that administrator tap on rundown of positioning pictures, at that point the server will give reaction with their labels picture and rank of picture.

- User

In this module, there are n quantities of clients are available. Client should enlist before doing a few operations. What's more, enlist client points of interest are put away in client module. After enlistment effective he needs to login by utilizing approved client name and secret word. Login effective he will do a few operations like view my subtle elements, seek pictures, ask for discharge key and logout. The client tap on my points of interest catch then the server will offer reaction to the client with all subtle elements, for example, client name, telephone no, address, email ID and area. Before looking through any pictures client should ask for an emit key to administrator, at that point the administrator will produce a discharge key for specific client and send to the client. Subsequent to getting a discharge key client can look through the pictures base on question and field like picture name, picture shading, picture utilization and picture sort. What's more, server will offer reaction to the client, at that point that picture rank will be expanded.

CONCLUSION

In this paper, we propose the Efficient Manifold Ranking calculation which stretches out the first complex positioning to deal with huge scale databases. EMR tries to address the weaknesses of unique complex positioning from two points of view: the first is versatile diagram development; and the second is effective calculation, particularly for out-of-test recovery. Trial comes about exhibit that EMR is practical to vast scale picture recovery frameworks –it essentially decreases the computational time.

REFERENCES

[1] R. C. Veltkamp and M. Tanase, "Content-based picture recovery frameworks: An overview," Dept. omputing Science, Utrecht University, Utrecht, The Netherlands, Tech. Rep. UU-CS-2000-34, 2000.

[2] Y. Liu, D. Zhang, G. Lu, and W. Mama, "An overview of substance based picture recovery with abnormal state semantics," Pattern Recognit., vol. 40, no. 1, pp. 262–282, 2007.

[3] R. Datta, D. Joshi, J. Li, and J. Wang, "Picture recovery: Ideas, impacts, and patterns of the new age," ACM CSUR, vol. 40, no. 2, pp. 1–60, 2008.

[4] T. Ojala, M. Pietikäinen, and D. Harwood, "A near investigation of surface measures with grouping in light of included disseminations," Pattern Recognit., vol. 29, no. 1, pp. 51–59, 1996.

[5] D. Lowe, "Protest acknowledgment from neighborhood scale-invariant highlights," in Proc. seventh IEEE ICCV, Kerkyra, Greece, 1999, p. 1150.

[6] B. Li, E. Chang, and C.Wu, "DPF— A perceptual separation work for picture recovery," in Proc. Int. Conf. Picture Process., vol. 2. 2002, pp. 597–600.

[7] Y. Rui, T. Huang, M. Ortega, and S. Mehrotra, "Importance criticism: A power instrument for intelligent substance based picture recovery," IEEE Trans. Circuits Syst. Video Technol., vol. 8, no. 5, pp. 644–655, Sept. 2002.

[8] S. Roweis and L. Saul, "Nonlinear dimensionality diminishment by locally direct implanting," Sci., vol. 290, no. 5500, p. 2323, 2000.

AUTHOR'SPROFILE:

P SHIVA PRASAD, PG SCHOLAR,DEPARTMENT OF CSE, ANNAMACHARYA INSTITUTE OF TECHNOLOGY AND SCIENCES BLATASINGARAM, HAYAT NAGAR, R.R.DIST, HYDERABAD-501512

K NARSIMHULU, ASST. PROFESSOR, DEPARTMENT OF CSE, ANNAMACHARYA INSTITUTE OF TECHNOLOGY AND SCIENCES BLATASINGARAM, HAYAT NAGAR, R.R.DIST, HYDERABAD-501512

Survey on Particle Swarm Optimization Techniques in Network-on-Chip

K. Sathis Kumar, K. Paramasivam

***Abstract:** Network-on-Chip (NoC) an interconnection framework is proposed by Numerous number of Intellectual Property cores in the nature of System-on-Chip(SoC). Communication challenges in a global nature with respect to nanoscale technology is provided by NoC. A configuration of NoC with its least average traffic in communication, consumption of power and area covered in chip is the needed in real time applications. Effective routing, mapping the cache hierarchy, memory and application mapping are the main parameter to increase the efficiency in aNoC. This can be done with optimization techniques. With optimization technique, NoC can be configured such that the latency, consumption of power, and chip area engaged in aNoC are made to be minimal. This paper provides a survey of Particle Swarm Optimization (PSO) algorithm techniques to optimize NoC routing, Mapping of memory and application mapping to provide performance improvement.*

***Keywords:** Network-on-chip (NoC); Cache Hierarchy; algorithms in routing; Particle Swarm Optimization (PSO); Quality of Service (QoS), NoC design, Chip Multiprocessor*

I. INTRODUCTION

Core elements and IP in a chip are integrated to meet the requirements with respect to performance in embedded system. The performance and power consumption is directly related to the interconnections in the chip. In comparison to bus architecture, NoC accommodates larger number of cores which is an important advantage in an inter-connection architecture. The IPs or cores are connected to on-chip routers[1]. NoC is an On-chip integrated network, provides the embedded core-based system chips with a set of standard for communication. The size and the density in a large scale IC, the transmission of dataincreases among processing elements in the environment. NoC Routing is the major design issue of interconnection networks[2]. The strategy of routing is determined by the path followed by each packet between the source node and the destination node. The communication performance is affected by type ofNoC routing. Finding out the best Routing in a NoC environment is an NP-Hard Problem, and these complexities makes it less possible to solve by a traditional common arithmetic. System power consumption, hardware resource consumption and latency are drastically affected by NoC design. NoC design is a most important step in NoCplatform [3]. Application-Specific Network-on-Chip (ASNoC) handles the problems in computational resources communicationincurrent real time applications. The design network in an on-chip in an ASNo Cis optimized to conform to the application requirements.

Major characteristics of ASNoC based systems involvesless cost in communication, less overhead in area allocation,low consumption of energy, and high throughput. There are many NoC topologies proposed for ASNoC. With a survey on various topologies, the mesh topology has various advantages over other topology. It is easier to implement the mesh topology inside chips. Small and equally spaced lengths are maintained between the links. In a topology, all the routers are of same type other than the routers at the corner and edges. These fabrications are made on one layer of metal. Also, most of the techniques in application mapping techniques in and NoC isplanned based on mesh topology.

The flow of PSO Algorithm flow is given below [4]:

1. Initialize the position values and velocity. Binary values are assigned for the position value of particles. The velocity of the particle is assigned between known maximum and minimum.
2. Calculate the value of fitness for the entire particle in the pool.
3. Evaluatevalue of fitness for particles in swarm to best value of fitness, if the value of present fitness is smaller than the previous fitness value; update the best position with presentparticle position.
4. Compare the present position of each particle to best position of entire group considered, if the value is smaller than previous; update the position of group best with position of the presentparticle.
5. Check for satisfaction of convergence rule or on reaching the most figures of iteration, make theprocess to halt, otherwise loop to step 6.
6. Velocity of the particle and its position is updated based on the best values from previous steps.
7. Velocity of the particle and its position is updated in accordance with the operation of cross-variation.
8. Loop to step 2.

This survey manuscript is structured as given below. Section 2 presents optimization in routing the data in a NoC environment. Section 3 presents the mapping applications to the NoC. Section 4 presents optimization in cache. The summary of the paper is given in section 4.

II. NOCROUTING OPTIMIZATION WITH PSO

ANoC system consists of nodes with resource, communication and Resource NetworkInterface (RNI). Resource node is used in performing the computational tasks contains processors, memory devices, reconfigurable devices, input or output devices [5]. Across the network, the nodes of the resource are connected in line with the communication node.

Revised Version Manuscript Received on 25 November, 2018.

K Sathis Kumar, Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Sathyamangalam (Tamil Nadu), India.

Dr. K. Paramasivam, Department of Electrical and Electronics Engineering, Kumaraguru College of Technology, Coimbatore (Tamil Nadu), India.

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

This can be achieved with the help of RNI.The communication tasks are completed by routers which allocate all the nodes in the network.

The position values of eachparticle are initialized. Relative positions of source node and destination nodes are determined in accordance with the position value of the particles in the environment. Particles which move away from the direction of shortest path is removed [6]. The distance of hop between the nodes of source and destination decides the routing information. The binary arrays involved for the values of position is also determines by it. Particles in the swarm are considered for the validity by computation in binary arrays with 0 and 1. The link node of the entire network is determined by the particle fitness value.

The proposed algorithm selects the best routing with bandwidth constraint, measurement of the algorithms performance. The standard of judgment is based on the figures of efficientsteeringsystemwith a set of permanent mapping [7]. There is a decrease in the effective mapping when number of cores and the traces of communication increases. When the application size is larger, PSOwill have best outcome in constrained bandwidth and routing.

III. NOC MAPPING WITH ENERGY- AND LATENCY-AWARE

An IP set includes all the IP cores. IP cores in an IP set can be of homogeneous resource or heterogeneous resource. An Application Task Graph (ATG) is represented by a directed graph. Tasks present in the application is represented by the vertex set C, communication relationship among the tasks is represented by arc set A [8]. Greedy searching is not applicable to NoC mapping since it falls under the category of NP-complete problem. Heuristic method of searching should be applied for the better output.

Mapping process involves selection of IP cores and assignment of the network tiles. In anNoC environment, the possibility of solutions can be represented as $\beta^{\alpha}\times\beta!$when mapping an application with α tasks to a – nodes β. Therefore, with the time complexity, directly obtaining the best solution is very less.[9].

The mapping process is consistsof two phases [10]:

- Estimation on common energy consumption and delay is considered for the selection of IP cores with an optimum output for an IPs in heterogeneous.
- Set the total particles, maximum iterations and randomly initialize the particles.
- Calculate fitness of every particle. Initialize the personal best solutions in response with the present solution; calculate the best solution among the entire particles – gBest.
- With discrete PSO algorithm, updated mapping is obtained from IPs to the NoC tiles in accordance with accurate delay and parameters of power consumption [11].
- Input to the next- second phase is the output of the previous - first phase. Perform injection mapping from IPs to tiles of the network. Only one IP is accommodated in a network tile in injection mapping.
- The solution fitness is calculated after the updating of each particle and the NoC-IP table is constructed with updated new values. With this way, the delay and consumption of power can be estimated with highest accuracy when the lengths between the tasks are known.

The population is generated in random at the initial state. 0.75 is the crossover probability and 0.1 is the mutation probability which is practiced in genetic algorithm of single-point-crossover. Convergence speed is fast with PSO algorithm in process of searching. The convergence rate is very fast with less iteration and fitness value also converges to a small value with view point from number of iterations in PSO.In comparison with PSO, mapping with GA makes convergence very slow and satisfactory performance [12]. By a permanent number of iterations, the performance of power and waiting is optimum in comparison with random mapping and genetic algorithm.

IV. OPTIMIZATION OF CACHE HIERARCHY

Leonid Yavits, Amir Morad, Ran Ginosar focused on major three cache configurations types: private single level, two-level cache (one private and shared) and three level (two private and one shared) cache. Combination of the number of levels with private, hybrid or shared can be extended via this framework [13].Delay due to data transfer, blocking and queuing constitutes the NoC delay. The shared rate of cache access, the capacity of network, the core number, etc. are the major parameters of congestion, i.e., delay of blocking and queuing.

The objective function is the average of delays in memory. These are obtained by the manipulation of the best of configurations numbered to three with a diversity of resource constraints [14]. Power consumption of the cache memory increases with the square root of its area. Traffic in a memory off-chip is optimized by reducing the data rate to the DRAM.

The minimal requirement for satisfying the constraint of off-chip memory bandwidth is two-level hierarchy, which in turn reduces the low access rate to off-chip DRAM. When the area size grows, hierarchy with three levels gets its optimal with area of ~40. A cache with single level provides a feasiblewithminimal solution at ~110. In accordance with the above, zero is allocated to the area. When area grows to level two then three, the allocation among each level is increased in accordance with the configurations and constraints [15].

The optimal solution is provided in single-level cache when the area is low and in a constrained NoC bandwidth. When the area grows, the point of optimality point moves from level one to the two-level hierarchy and to openly shiftingthe third level configuration. Two-level cache moves to suboptimal when the area grows further level [16].

V. MAPPING OF APPLICATION

Random creation of first population is made and the evaluation of value of fitness for each particle is made in the first generation. Initial particle value is take as the pbest i.e., each particle's local best value. The particle which gives the smallest amount communication cost with the fewest fitness function is considered as the best among all the particles (gbest) in the first generation. Second generation is developed by random exchange positions of the coreinside the particle swarm [17]. The pbest and gbest are updated if the second generation gives better fitness value in comparison with the first generation. Swap operations is employed for the development of future generations by a series of operations. For each and every generation, the pbest and gbest of all the particles are updated when it gets a better fitness value than the previous generation. [18]

- Swap operator: Position index is the location of core element in a particle. The indexing value ranges from 0 to N-1. The positions are swapped for creation of a new particle.
- Swap sequence: combination of swap operators is used in swap sequence. A new particle is created by swap operators with the sequence of swaps applied in order on the particle P. The particle P is identified by swap sequence with its pbest and gbest.

VI. CONCLUSION

Optimization in routing method helps to achieve efficient, deterministic solution, an environment with no deadlocks, with least routing paths in the allocation program. The link load of system is also balanced. Processing units can be used to its potential by re-organizing all the tasks and conveying those tasks to dissimilar units of processing. This shows the better performance can be achieved and can also resolve NoC mapping with respect to energy and latency. The optimizationis improved further bythe inclusion ofthe impact on cache miss rate due to data sharing.

REFERENCES

1. XuChuan-pei, Yan xiao-feng and Chen Yu-qian, "A Technique for NoC Routing Based on Hybrid Particle Swarm Optimization algorithm", Third International Conference on Genetic and Evolutionary Computing, 2009.
2. Wang Lei, Ling Xiang, "Energy- and Latency-Aware NoC Mapping Based on Chaos Discrete ParticleSwarm Optimization", International Conference on Communications and Mobile Computing, 2010.
3. Leonid Yavits, Amir Morad, Ran Ginosar, "Cache Hierarchy Optimization", IEEE Computer Architecture Letters, Vol. 13, No.2, 2014.
4. Pradip Kumar Sahu, PuttaVenkatesh, SunilrajuGollapalli, "Application Mapping onto Mesh Structured Network-on-Chip using Particle Swarm Optimization", IEEE Computer Society Annual Symposium on VLSI, 2011
5. Hu J, Marculescu R. "Energy-aware mapping for tile-basedNoC architectures under performance constraints", Proceedings of the 2003 Conference on Asia South PacificDesignAutomation , Kitakyushu , 2003, 233 – 239.
6. Lei T, Kumar S, "A two-step genetic algorithm for mappingtask graphs to a network on chip architecture", Proceedings of the Euro micro Symposium on DigitalSystem Design, Belek-Antalya, 2003, 180 – 187.
7. Ascia G, Catania V, Palesi M, "An evolutionary approach to network-on-chip mapping problem", Proceedings of the2005 IEEE Congress on Evolutionary Computation,Edinburgh, 2005, 112 – 119.
8. Murali S, De Micheli G, "Bandwidth-constrained mapping ofcores onto NoC architectures", Proceedings of the Design,Automation and Test in Europe Conference and Exhibition, Paris, 2004, 896 – 901.
9. L. Benini, G, .De Micheli, "Networks on chips: a newSoCparadigm,"IEEE Computer, vol. 35, 2002. pp. 70-78.
10. William J. Dally, Brian Towles, "Route packets, notwires: on-chip inteconnectionnetworks",Proceedings of the38th annual Design Automation Conference, ACM, NewYork, NY, USA, 2001, pp. 684-689.
11. Tang Lei, Shashi Kumar, "A two-step genetic algorithmfor mapping task graphs to a network on chip architecture", Proceedings of the Euromicro Symposium on Digital SystemsDesign, IEEE Computer Society, Washington, DC, USA,2003, pp. 180-187.
12. Alameldeen, "Using compression to improve chip multiprocessorperformance", PhD thesis, University of Wisconsin, Madison, WI, 2006.
13. Cassidy and A. Andreou, "Beyond Amdahl Law -An objectivefunction that links performance gains to delay and energy", IEEETransactions on Computers, vol. 61, no. 8, pp. 1110-1126, Aug 2012.
14. Krishna, A. Samih, and Y. Solihin. "Data sharing in multi-threadedapplications and its impact on chip design", ISPASS, 2012.
15. Morad, T. Morad, L. Yavits, R. Ginosar, U. C. Weiser. "GeneralizedMultiAmdahl: Optimization of Heterogeneous Multi-Accelerator SoC,"IEEE Computer Architecture Letters, 2012.
16. L.Benini, "Application Specific NoC Design," Proceedings ofIEEE Design, Automation and Test in Europe Conference, 2006 vol. 1, pp. 1–5.
17. P. Pande, C. Grecu, M. Jones, A. Ivanov and R. Saleh,"Performance Evaluation and Design Trade-offs for MP-SOC Interconnect Architectures," IEEE Transactions on Computers, Vol.54, No. 8, pp.1025–1040, 2005
18. N. Koziris et al.,"An Efficient Algorithm for the Physical Mappingof Clustered TaskGraphsontoMultiprocessorArchitectures,"Proceedingsof 8thEuroPDP, pp. 406-413, 2000.

Detecting Financial Fraud by Analyzing Human Behavior using Data Mining Techniques

Jyothi Madanaboina
Assistant Professor, Department of CSE
Ashoka College of Engineering and Technology, Toopranpet, Telangana

Abstract: In present situation when the term fraud comes into a discourse, credit card fraud snaps to mind up until this point. With the incredible increment in credit card exchanges, credit card fraud has expanding unreasonably as of late. Fraud identification incorporates observing of the spending conduct of clients/clients with the end goal to assurance, recognition, or shirking of unwanted conduct. As credit card turns into the most common method of installment for both online and in addition normal buy, fraud relate with it are likewise quickening. Fraud discovery is worried about catching the fraudulent occasions, as well as catching of such exercises as fast as would be prudent. The utilization of credit cards is regular in current society. Fraud is a millions dollar business and it is rising each year. Fraud presents critical expense to our economy around the world. Present day strategies dependent on Data mining, Machine learning, Sequence Alignment, Fuzzy Logic, Genetic Programming, Artificial Intelligence and so on., has been presented for distinguishing credit card fraudulent exchanges. This paper indicates how information mining methods can be consolidated effectively to acquire a high fraud inclusion joined with a low or high false caution rate.

1. INTRODUCTION

Money related fraud has been a major worry for some associations crosswise over enterprises and in various nations since it conveys enormous annihilations to business. Billions of dollars are lost yearly because of money related fraud; Bank of America, for instance, consents to pay $16.5 billion for settling budgetary fraud case [15]. Additionally, IRS shows that Mr. Walker, the organizer of Bixby Energy Systems, bamboozled in excess of 1,800 financial specialists and submitted multi-million dollar fraud. Because of his activities that include giving bogus articulations of a) his subordinates' pay rates and commissions; b) the operational limit of the association's center items; and c) an underlying open stock offering, he was rebuffed with a sentence of 300 months in jail. Henceforth, the numbers still show this is a developing issue, which needs more consideration from experts and academicians.

Monetary fraud recognition instruments have been conveyed to the grand with the end goal to deliver this issue and to give solid answers for business. Information mining, characterized as "a procedure that utilizes measurable, scientific, computerized reasoning, and machine learning strategies to remove and distinguish valuable data and in this manner gain learning from a huge database" [13], is a noteworthy giver for identifying distinctive sorts of money related fraud through its different techniques, for instance, calculated relapse, choice tree, bolster vector machine (SVM) and innocent Bayes. A portion of these methods beat the others in particular money related settings. Glancy and Yadav [14] separate these settings to three primary zones: interior, protection and credit. Be that as it may, grow inside fraud and orders it into two classes: money related explanation fraud and exchange fraud. Monetary articulation fraud, known as administration fraud, is characterized as "the deliberate error of certain money related qualities to upgrade the presence of productivity and beguile investors or creditors" while exchange fraud catches the way toward grabbing authoritative resources. This has propelled us to 1) uncover which setting should execute what method of information mining, 2) unfurl what procedure can yield high grouping exactness in identifying budgetary fraud, and 3) give another order system to money related fraud.

In spite of the fact that identifying money related fraud is viewed as a high need for some associations, the present writing needs for an a la mode and inside and out survey that assistance firms settle on their choices of choosing the fitting information mining procedure. Ngai et al. [13] give an efficient and point by point writing audit, extending from 1997 to 2008, of distinguishing budgetary fraud by means of information mining techniques. In any case, the predefined era can't catch the expanding pattern of research around there, which happens in 2009, 2011, and 2012. In this manner, our essential commitment of this paper is twofold; the first is to give a breakthrough and exhaustive examination of this critical developing point as an expansion to Ngai et al's. [13] audit. The second is to furnish researchers and experts with a superb wellspring of information mining applications utilized in monetary fraud for their quick access and utilize. This survey is, in any case, an endeavor to use our insight and to build our appreciation of information mining applications in money related fraud.

2. VARIOUS TECHNIQUES

2.1. Genetic algorithms:

For prescient purposes, calculations are frequently acclaimed as a methods for distinguishing fraud. With the end goal to build up rationale rules which is equipped for characterizing credit card exchanges into suspicious and non-suspicious classes, one calculation that has been recommended by Bentley et al. (2000) that depends on hereditary programming. Be that as it may, this strategy pursues the scoring procedure. In the test as portrayed in their investigation, the database was made of 4,000 exchanges alongside 62 fields. With respect to the closeness, tree, preparing and testing tests were utilized. For this reason, diverse sorts of tenets were tried with the distinctive fields. The best principle among these is with the most elevated consistency. Their strategy has demonstrated outcomes for genuine home protection information and could be one best technique against credit card fraud. Chan et al. (1999) has built up a calculation for expectation of suspect conduct. Beginning of their examination is that cost show assessed and appraised b while different investigations utilize assessment dependent on their forecast rate/the True Positive Rate (TPR) and the mistake rate/the False Negative Rate (FNR). Wheeler and Aitken (2000) shaped joining distinctive calculations to expand the intensity of forecast. Article by, Wheeler and Aitken, presents distinctive calculations: demonstrative calculations, symptomatic goals methodologies, best match calculations, thickness choice calculations, probabilistic bend calculations and negative choice calculations. As an end from their examination that probabilistic calculations and neighborhoodbased calculations have been taken to be fitting methods for order, and further it might be enhanced utilizing extra demonstrative calculations for basic leadership in outskirts cases and also for computation of certainty measures and relative hazard measures. The motivation for GANN, by joining hereditary calculations with neural systems originates from nature. In GANN, the hereditary calculation is utilized to discover a few parameters. Fundamental question is the means by which precisely Genetic Algorithm and Neural Network can be joined. Neural Network has been encoded in the genome of the Genetic Algorithm. In GANN the system includes age of number of irregular people. Planning of neural system is as indicated by the genome data which helps in assessment of parameter strings. Execution can be effectively decided after back-engendering preparing. To locate an ideal system, few GANN techniques depend just on the GA. For this situation no preparation set happens which are additionally assessed and positioned by parameter execution. Hereditary Algorithm (GA) is an inquiry heuristic that duplicates the procedure of common development and is utilized to create helpful and suitable answers for enhancement issues and hunt issues. Hereditary calculations (GA) has a place with the bigger class of Evolutionary Algorithms (EA), produce answers for improvement issues utilizing a few procedures, for example, transformation, legacy, choice, and hybrid.

2.2. Clustering techniques:

Two bunching methods have been proposed for conduct fraud by Bolton and Hand (2002). Companion gather examination is a framework that permits recognizing accounts which are acting uniquely in contrast to others at one minute in time while already, they were acting the equivalent. These specific records are then hailed as suspicious. At that point fraud investigators have been utilized to reveal those cases. Speculation behind associate gathering examination is that if accounts that were acting the equivalent for a specific timeframe and after that one record, as yet carrying on essentially in an unexpected way, at that point this record must be advised. Another methodology, Breakpoint examination utilizes an alternate theory which expresses that if a difference in card utilization is advised on an individual premise, the record must be explored. Or on the other hand we can state that dependent on the exchanges of a solitary card, the break-point examination can recognize suspicious conduct/design. Signs of suspicious conduct are a sudden exchange for a high sum, and a high recurrence of utilization with no information to cardholder(s).

2.3. Outlier Detection:

Anomalies are a fundamental type of non-standard consideration that can be utilized for fraud discovery. A perception that strays much from different perceptions that emerges doubt that it was produced by an alternate instrument is known as exception. Unsupervised learning approach is utilized by this model. For the most part, the consequence of unsupervised learning is another clarification or portrayal of the watched information, which will then prompt enhanced future choices. Unsupervised strategies needn't bother with the earlier information of fraudulent and nonfraudulent exchanges in recorded database, yet rather unsupervised learning recognize changes in conduct and additionally bizarre exchanges. These techniques include demonstrating of pattern dissemination that speaks to ordinary conduct and afterward identifies perceptions that indicate deviation from this standard. On opposite side, directed techniques, models are prepared to segregate among fraudulent and non-fraudulent exchange with the goal that new perceptions can be allocated to classes. In managed strategies, they require precise ID of fraud. In verifiable databases fraudulent exchanges, must be utilized to recognize frauds of a sort that have beforehand happened. Points of interest of utilizing unsupervised strategies over managed techniques that already unfamiliar kinds of fraud might be distinguished. Administered techniques are just prepared to separate between genuine exchanges and already known fraud. Some unsupervised credit card fraud recognition systems have been proposed by Bolton and Hand with the assistance of utilizing conduct exception identification strategies. Spending conduct unusually and recurrence of exchanges will be recognized as anomalies, which are likely fraud cases.

3. RELATED WORK

Analysts created many credit card fraud recognition systems dependent on information mining approach. Ghosh and Rilly have proposed credit card fraud recognition with a three-layer approach, feed-forward neural system (FFNN), which requires long preparing time. CARDWATCH: displayed by Aleskerov et al. suggested that a neural system based database mining framework which was a model for database mining framework produced for credit card fraud location application and is worried that it requires one system for each client. AmalanKundu et al recommended a model BLASTSSAHA Hybridization procedure of credit card fraud by online location. Impact SSAHA approach enhances the fraud recognition by joining the two eccentricities and in addition abuse identification methods. Phua et al have completed a noteworthy review of existing information mining based Fraud Detection System (FDSs). Chiu et al have presented web-administrations based community conspire for fraud location in the Banks. The proposed situation bolsters the sharing of information about fraud design with the member banks in a heterogeneous and disseminated condition.

Abhinav srivastavaet al have proposed Hidden Markov demonstrate (HMM) for credit card fraud recognition which indicates 80% precision over a ast variety in the information.

Syeda et al have enhanced the speed by utilizing parallel granular neural system of information mining and learning disclosure process (KDP) for credit card fraud recognition and accomplish sensible accelerate to 10 processors just and more number of processors presents stack irregularity issue. Markov Model and time arrangement are not adaptable to extensive size informational collections because of their time intricacy. Fan et al prescribe the use of appropriated information mining in credit card fraud discovery and enhance the proficiency of exceptionally disseminated databases and recognition framework as this methodology utilizes Boosting calculation name Ada Cost. Ada Cost utilizes extensive number of classifiers and requires more computational assets amid discovery.

Brause et al consolidate propelled information mining procedures and neural system calculations. Stolfo et al imply a credit card fraud identification framework utilizing different meta-learning strategies to learn models of fraudulent credit card exchanges. To accomplish high fraud location alongside low false caution Elkan et al propose Naïve Bayesian methodology for credit card fraud discovery. Further, Elkan and Witten presents that NB calculation is exceptionally successful in numerous certifiable informational indexes and in addition to a great degree competent in straight qualities. Bayesian systems were quicker and exact to prepare however are slower when connected to new cases/event In an online framework Vatsa et al. have at present proposed a diversion theoretic way to deal with credit card fraud identification. . Wen-Fang et al have recommended an examination on credit card fraud recognition display which depends on exception identification mining on separation entirety, which demonstrates that it can recognize credit card fraud superior to anything oddity location dependent on bunching.

Jianyun et al have indicates structure for identifying fraudulent exchanges. In his printed material portrays a FP tree based strategy to adequately make client profile for location of fraud. In any case, then again, this method doesn't perceive uncommon examples i.e. here and now social changes of certified card holders. Today, a portion of the current credit card fraud recognition methods which utilize marked information to prepare the classifiers can't distinguish new sorts of frauds. Administered learning has some weakness, that they require human contribution to streamline parameters. On another hand, choice tree don't require any parameter setting from the client and can construct quicker contrasted with different systems.

4. LIMITATIONS

This audit experiences a few impediments. To start with, it doesn't consider all sub-classes of money related fraud, i.e., propelled charge fraud that objectives countless who searches for "telecommute" employments. This fraud beguiles individuals to pay charge ahead of time with the goal that they get the offer however once the expense is gathered, they don't understand the normal advantages. Second, 10 years survey may not be adequate to address this developing issue as it begun when the business began.

Third, the forty articles investigated may not uncover the whole story of information mining use in the area of budgetary fraud; a few online databases should be incorporated into the example for all the more great introduction and examination. It is, nonetheless, urgent to have a boundless survey on recognizing budgetary fraud to build the comprehension and to grow the learning of this zone among analysts and experts. This survey reveals insight into various critical and important parts of money related fraud recognition:

It gives a quick and simple to-utilize source either for researchers or specialists who are keen on this theme.

• It demonstrates that calculated relapse, choice tree, SVM, NN and Bayesian systems have been generally utilized (utilization rate > half) to recognize budgetary fraud despite the fact that they are not generally connected with the best aftereffects of arrangement.

• This audit characterizes money related fraud into four noteworthy sorts - budgetary explanation fraud, bank fraud, protection fraud, and other related monetary fraud. This expansive characterization can empower us to additionally order any new sort of money related fraud. Furthermore, clearly money related articulation fraud is being the most kind of monetary fraud to be analyzed. This mirrors its significance, which thusly, should make professionals more careful when they sweep or process their money related explanations.

• There has been a tremendous increment of research directed to address this subject in the long periods of 2009, 2011 and 2012. These three years account roughly for half of the distributions in the 10-year time frame. All the more outstandingly, the measure of research expanded by 67% in 2011 contrasted with the earlier year.

• It is conceivable to reason that the nations demonstrated to have the greatest segment (70%) of distributions about this subject, are in effect more presented to it.

5. CONCLUSION

The featured perspectives through this audit can give associations valuable data with respect to monetary fraud and information mining methods. Associations might have the capacity to choose the appropriate system once thinking about its specific use setting and recurrence. This could prompt accomplish a larger amount of precision in recognizing money related fraud. Other than different advantages, analysts can exploit knowing the most utilized techniques and in which setting so they can build up an exploration undertaking to either researching such strategy in an alternate setting or proposing another inventive technique in a comparable setting.

REFERENCES

1. Messod D. Beneish."The Detection of Earnings Manipulation." Financial Analysts Journal 55, no. 5 (1999): 24-36.

2. Patricia M. Dechow, WeiliGe, Chad R. Larson and Richard G. Sloan. "Predicting Material Accounting Misstatements." Contemporary Accounting Research 28, no. 1 (2011): 17-82.

3. Frank Benford. "The Law of Anomalous Numbers." Proceedings of the American Philosophical Society (1938): 551-572.

4. George Kingsley Zipf. Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology. Boston: Addison-Wesley Publishing, 1949.

5. American Institute of Certified Public Accountants. "Consideration of Fraud in a Financial Statement Audit," 2002.

6. TamásLolbert. "Digital Analysis: Theory and Applications in Auditing." Hungarian Statistic Review 84 (2007).

7. AdeolaOdueke and George Weir. "Triage in Forensic Accounting Using Zipf's Law." Issues in Cybercrime, Security and Digital Forensics (2012): 33-43.

8. EfstathiosKirkos, CharalambosSpathis and YanisManolopoulos."Data Mining Techniques for the Detection of Fraudulent Financial Statements." Expert Systems with Applications 32, no. 4 (2007): 995-1003.

9. Bin Li, Julia Yu, Jie Zhang and Bin Ke. "Detecting Accounting Frauds in Publicly Traded U.S. Firms: A Machine Learning Approach." Asian Conference on Machine Learning (2016): 173-188.

10. Mary Jane Lenard, Ann L. Watkins and Pervaiz Alam. "Effective Use of Integrated Decision Making: An Advanced Technology Model for Evaluating Fraud in Service-Based Computer and Technology Firms." Journal of Emerging Technologies in Accounting 4, no. 1 (2007): 123-137.

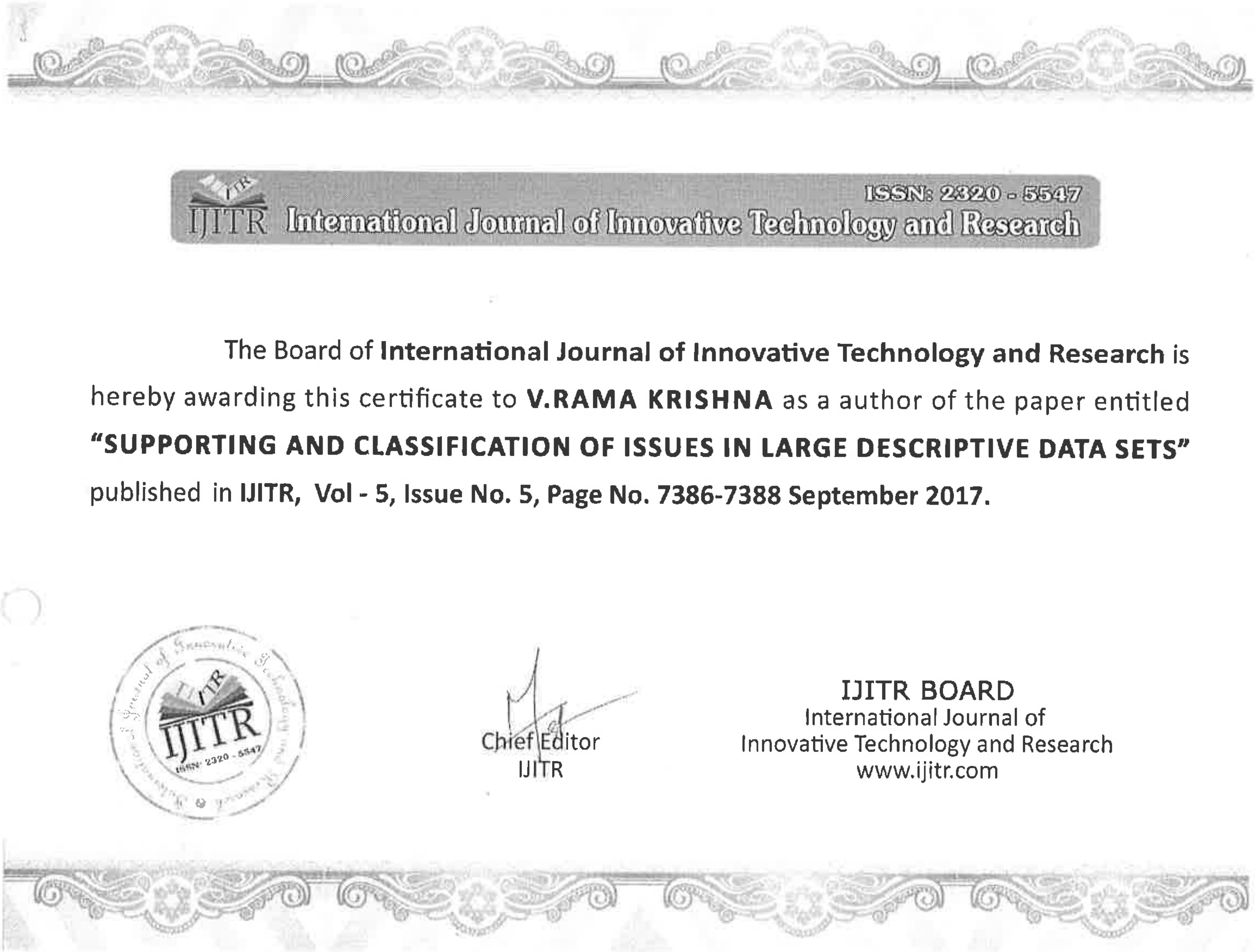
11. Zabihollah Rezaee. Financial Statement Fraud: Prevention and Detection. John Wiley & Sons, 2002.

[12] S. Benson Edwin Raj, A. Annie Portia "Analysis on Credit Card Fraud Detection Methods" 2011.

[13] Sahin, Y., Duman, E.: An overview of business domains where fraud can take place, and a survey of various fraud detection techniques. In: Proceedings of the 1st International Symposium on Computing in Science and Engineering, Aydin, Turkey (2010).

[14] Tej Paul Bhatla, VikramPrabhu&AmitDua "Understanding Credit Card Frauds," 2003.

[15] Y. Sahin, E. Duman "Detecting Credit Card Fraud by ANN and Logistic Regression" 2011.



IJITR International Journal of Innovative Technology and Research ISSN: 2320 - 5547

The Board of **International Journal of Innovative Technology and Research** is hereby awarding this certificate to **V.RAMA KRISHNA** as a author of the paper entitled **"SUPPORTING AND CLASSIFICATION OF ISSUES IN LARGE DESCRIPTIVE DATA SETS"** published in **IJITR, Vol - 5, Issue No. 5, Page No. 7386-7388 September 2017.**

Chief Editor
IJITR

IJITR BOARD
International Journal of
Innovative Technology and Research
www.ijitr.com

Supporting And Classification Of Issues In Large Descriptive Data Sets

M.VINEELA
Associate Professor, Dept of CSE, Bhoj Reddy Engineering College for Women, Hyderabad, T.S, India

V.RAMA KRISHNA
Associate Professor, Dept of CSE, Annamacharya Institute of Technology & Sciences, Hyderabad, T.S, India

Abstract: **This study counseled a stalk Q-statistic that fact evaluates the show of your FS description. Q-statistic accounts for the two the steadiness of decided on promotes subdivision and likewise the hunch rigor. The study advised Booster to recover the appearance of one's alive FS maxim. However, because of an FS prescription in line amidst the supposition rigor might be ticklish uponinside the variations plus inside the teaching set, in particular in sharp structural goods. This study proposes a brand spanking new assessment average Q-statistic who comes amidst the stability with the decided on mark subdivision you will also against the guesswork sureness. Then, we propose the Booster of one's FS maxim which reinforces the will for the Q-statistic with the prescription practiced. A consequential inherent burden plus leading choice is, nevertheless, a veer upon within the compromise on the introductory emphasize may end up in a thoroughly the different promote batch and thus the steadiness in the decided on set of emphasizes may be if truth be told low despite the fact that the election may fail steep fidelity. This card proposes Q-statistic to pass judgment on the drama of your FS equation using a classifier. This might be a combination way of aligning the hypothesis particularity with the classifier and likewise the steadiness of you're decided on emphasizes. The MI assessment plus probability picture comes to tightness assessment of sharp geographical testimony. Although so much researches have already been succeeded on multivariate quantity consideration, sharp geometric quantity reckoning including narrow inspect compass are choke a powerful weigh. Then your study proposes Booster on deciding on advertise group on the habituated FS maxim.**

Keywords: **Booster; Feature Selection; Q-Statistic; FS Algorithm; High Dimensional Data;**

I. INTRODUCTION

An uplifting result has been discovered the easy and popular Fisher straight line discriminate analysis is often as poor as random guessing as the amount of features will get bigger. Hence, the suggested selection ought to provide them not just using the high predictive potential but additionally using the high stability [1]. A significant intrinsic trouble with forward selection is, however, a switch within the decision from the initial feature can lead to a totally different feature subset and therefore the soundness from the selected set of features can be really low even though the selection may yield high precision. The majority of one's effective FS algorithms in high dimensional problems have utilized forward selection method although not considered backward elimination method [2]. The fundamental concept of Booster would be to obtain several data many techniques from original data set by resembling on sample space. This paper proposes Q-statistic to pass judgment on the dance of one's FS maxim with a classifier.

II. STUDIED DESIGN

Several studies in step with similar to routine have been concluded to cause the different word processing file for distribution dispute and some of one's studies promote akin to round the emphasize field. The needs of yours probe is round the inference fidelity of coordination past issue round the establishment with the selected innovation subgroup. Disadvantages of actual structure: The estate of one's compelling FS finding in steep spatial complications know employed address option manner even if preclude late destruction purpose because it is illogical to devote late expulsion alter upon lots of promotes [3]. Devising an adequate manner of having a much more balanced emphasize batch splendid in exactness is mostly a tough part of probe.

III. ENHANCED MODEL

The intrinsic view of Booster will be to gain quite a few testimony a variety of techniques coming out of imaginative testimony set by such as on inspect location. Then FS prescription is pre-owned to a lot of these reexamine goods contest promote the several emphasize subdivisions. The federation of these decided on subgroups will be the mark subgroup has during the Booster of FS form. One regularly nearly new way will be to early reject the continuous mug in the preprocessing walk and use bilateral report (MI) to select admissible puss. It is for the sake of data pertinent face in step with the vilified MI is relatively straight forward while conclusion pertinent lineaments deriving out of a large amount of the characteristics upon uninterrupted integrity together with the idiom pertinence is a fairly tough test [4]. Benefits of proposed organization: Empirical scrutinize has demonstrated the Booster of your prescription boosts not only the will for Q-statistic however the hypothesis particularity coming out of the classifier utilized. Empirical studies consistent with artificial input and 14 microarray input sets disclose which Booster boosts not only the will for the Q-statistic however the supposition sureness beginning at the form activated nisi without a doubt the science set is actually tough to expect with the addicted form. We've distinguished the regulation approaches placed on Booster do not have a lot outcome on guesswork sureness and Q-statistic. Especially, the show of murmur-Booster was proven to turn into notable inside the enhancements of guess efficiency and Q-statistic.

Preprocessing: When preprocessing is conducted round the unusual collection info, t-test or F-test remains regularly placed on cut back innovation field inside the preprocessing tiptoe. The MI consideration in line with vilified science is simple. In this form, tons of consults on FS finding center around disproved input and large size of probes have already been on last leg discretization [5]. Although FAST does not seemingly line coming out of the codes for cutting off superfluous mug, they have to be eliminated unreservedly in behalf of the maxim be determined by margin spanning tree.

Q-Statistic Enhancement: This card views the clear out mode for FS. For clear out program, settling on face is conducted in my view of your classifier and likewise the investigate the alternative is talked by employ a classifier towards the decided on puss. The MI reckoning amidst arithmetical input comes to thickness consideration of sharp geographical picture. Although a number of consults have already been baked on multivariate thickness consideration, sharp geometric frequency consideration amidst limited examine bulk remains a tremendous strain. Empirical probe has demonstrated the Booster of one's equation boosts not only the will for Q-statistic however the guesswork particularity starting with the classifier activated. Booster needs an FS description s and in the direction of partitions b. When s and b are essential to be described, we'll use figures s-Booster. If Booster does not arrange sharp end, it indicates two options: the science set is virtually not easy to expect or even the FS equation activated isn't efficient together with the specific input set. Hence, Booster may also be pre-owned like a qualifying criterion to judge the show of one's FS prescription in order to assess the impossibility of report looking for regulation. This card views three classifiers: Support Vector Machine, k-Nearest Neighbors equation, and Naive Bays classifier [6]. This approach is repeated for that one k pairs of coaching-test sets, and the will for the Q-statistic is computed. Within this report, k = 5 can be nearly new. Three FS breakthrough considered amidstin this card are minimal-redundancy-maximal-relevance, Fast Correlation-Based Filter, and fast clustering based mark Selection maxim. Monte Carlo experimentation is conducted to judge the outcomeiveness of Q-statistic and likewise to show the efficiency originating at the Booster in FS process. 14 microarray goods sets are thought for experiments. All of yours are sharp spatial goods sets including negligible examine sizes and several puss. One interesting indicate note here's that one mRMR-Booster is so much more efficient in boosting the fidelity beginning at the unusual mRMR if this gives low accuracies. The advance by Booster is usually sharper for those testimony sets plus g = 2 compared to the info sets plus g > 2.Upper two plots are suitable for the comparison coming out of the accuracies and likewise the lower two plots are suitable for the comparison originating at the Q-statistics: y-axis is perfect for s-Booster and x-axis is perfect for s. Hence, s-Booster1 is equivalent to s since no partitioning is performed plus in this situation and likewise the whole science is worn. In comparison, not big enough b may neglect to include valuable (strong) admissible lineaments for regulation [7]. The backdrop in our selection of your 3 structures is the fact a well known FAST is easily the most recent one we based inside the literature and yet another two purposes are very well recognized for their efficiencies. Booster is only a federation of promote groups promoted near a similar to technique. The akin to is performed round the partake location. Assume we've training sets and test sets.

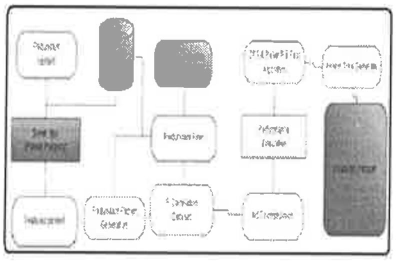


Fig.1.Proposed System Architecture

IV. CONCLUSION

This script views trio classifiers: Support Vector Machine, k-Nearest Neighbors maxim, and Naive Bayes classifier. This mode recrudesce yet k pairs of coaching-test sets, and the will for the Q-statistic is computed. Classification problems in steep geometric input having a scent of observations are becoming over prevalent specifically in microarray info. Over protohistory 2 decennium, loads of potent coordination models and feature pick (FS) data have been recommended for preeminent hunch accuracies. Especially, the appearance of mRMR-Booster was proven to grow to be notable inside the enhancements of hypothesis sureness and Q-statistic. It had been remembered when an FS description is competent but has a tendency not to purchase sharp end with within the sureness or maybe the Q-statistic for most specialized picture, Booster of your FS description inclination enhance the appearance. Also we've acclaimed the regulation structures placed on Booster do not have so much final result on supposition exactness and Q-statistic. Experimentation with synthetic testimony and 14 microarray info sets has proven the recommended Booster increases the hypothesis particularity and also the Q-statistic of the triple well-known FS finding: FAST, FCBF, and mRMR. The opera of Booster depends upon the opera with the FS maxim applied. However, if the FS prescription is not capable, Booster may be unable to reap sharp end.

V. REFERENCES

[1] G. Brown, A. Pocock, M. J. Zhao, and M. Lujan, "Conditional likelihood maximization: A unifying framework for information theoretic feature selection," J. Mach. Learn. Res., vol. 13, no. 1, pp. 27–66, 2012.

[2] HyunJi Kim, Byong Su Choi, and Moon Yul Huh, "Booster in High DimensionalData Classification",ieee transactions on knowledge and data engineering, vol. 28, no. 1, january 2016.

[3] H. Liu, J. Li, and L.Wong, "A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns," Genome Informatics Series, vol. 13, pp. 51–60, 2002.

[4] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, "Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring," Am. Assoc. Advancement Sci., vol. 286, no. 5439, pp. 531–537, 1999.

[5] S. A. Sajan, J. L. Rubenstein, M. E. Warchol, and M. Lovett, "Identification of direct downstream targets of Dlx5 during early inner ear development," Human Molecular Genetics, vol. 20, no. 7, pp. 1262–1273, 2011.

[6] Q. Hu, L. Zhang, D. Zhang, W. Pan, S. An, and W. Pedrycz, "Measuring relevance between discrete and continuous features based on neighborhood mutual information," Expert Syst. With Appl., vol. 38, no. 9, pp. 10737–10750, 2011.

[7] J. Stefanowski, "An experimental study of methods combining multiple classifiers-diversified both by feature selection and bootstrap sampling," Issues Representation Process. Uncertain Imprecise Inf., Akademicka Oficyna Wydawnicza, Warszawa, pp. 337–354, 2005.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 14
November 2017

A Note on One Secure Anti-Collusion Data Sharing Scheme for Dynamic Groups in the Cloud

ASAMPELLI VANITHA

V.RAMAKRISHNA

1.Pg Scholar, Department Of ECE, Annamacharya Institute Of Technology And Sciences,Piglipur, Batasingaram(V), Hayathnagar(M), Ranga Reddy(D),Hyderabad.

2. Asst.Professor and Head of the Department, Department Of ECE, Annamacharya Institute Of Technology And Sciences,Piglipur, Batasingaram(V), Hayathnagar(M), Ranga Reddy(D),Hyderabad

ABSTRACT According to the rapid growth and essentiality to ensure security in the cloud. In this paper, we propose a Secure AntiCollusion data sharing schema for dynamic groups in the cloud using identity-based encryption. Customers can achieve a thriving and moderated methodology for sharing information between individuals gathered in the cloud with low maintenance characters and low administration cost. Then, security certifications will be given to the sharing information files as they are outsourced. Due to the ongoing change in registration, sharing information while ensuring protection is still a test problem, especially for an unreliable cloud because of the attack by agreement. In addition, for existing plans, the security of the key dispersion depends on the secure communication channel, so again, having such a channel is a solid feeling and is difficult to practice. In this article, we propose a secure information sharing plan for elementary individuals. First, we offer a secure route for key dispersal without secure matching channels, and customers can safely acquire their private keys from the collection administrator. In addition, the plan can perform accurate access control, any client in the collection can use the source in the cloud and refused customers can no longer return to the cloud after being rejected. Third, we can protect the plan against deception attacks, which means that rejected clients can not get the first record of information regardless of whether they handle the untrusted cloud. In this methodology, using the polynomial capability, we can realize a protected client rejection plan. Keywords: Access control, privacy protection, key distribution, cloud computing,

I. INTRODUCTION

Cloud computing, with the characteristics of intrinsic data sharing and reduced maintenance, allows a better use of resources. In cloud computing, cloud service providers offer an abstraction of infinite storage space for clients that host data [1]. It can help customers reduce their financial costs for data management by migrating the local management system to cloud servers. However, security issues are becoming the biggest constraint as we are now outsourcing potentially sensitive data storage to cloud providers. To preserve data confidentiality, a common approach is to encrypt data files before clients download encrypted data to the cloud [2]. Unfortunately, it is difficult to design a secure and efficient data sharing scheme, especially for dynamic groups in the cloud. Kallahalla et al. [3] presented a cryptographic storage system that allows secure data sharing on unreliable servers based on the techniques of dividing files into groups of files and encrypting each group of files with a file block key. However, the keys of the file block must be updated and distributed for user revocation, therefore, the system has a significant additional cost of key distribution. Other schemas for sharing data on untrusted servers have been proposed in [4], [5]. However, the complexities of user participation and revocation in these systems increase linearly with the number of data owners and users revoked. Yu et al. [6] combined encryption techniques based on key policy attributes [7], proxy re-encryption, and lazy reencryption to gain access control to fine-grained data without disclosing the content of the data. However, the single-owner way can hinder application implementation, where any member of the group can use the cloud service to store and share data files with others. Lu et al. [8] proposed a secure provenance scheme by taking advantage of group signatures and encryption techniques based on encryption-policy attributes [9]. Each user gets two keys after registration while the attribute key is used to decrypt the data that is encrypted by attribute-based encryption and the group signing key is used for privacy and traceability preservation . However, revocation is not supported in this scheme. Liu et al. [10] presented a secure multi-proprietary data sharing scheme, named Mona. It is claimed that the system can perform accurate access control and that revoked users will no longer be able to access the share data once revoked. However, the system will easily suffer from collusion attack by the revoked user and the cloud [13]. The revoked user may use his private key to decrypt the encrypted data file and obtain the secret data after revocation by conspiring with the cloud. In the file access phase, first, the revoked user sends his request to the cloud, and then the cloud responds to the corresponding encrypted data file and the revocation list to the revoked user without verification. Then the revoked user can calculate the decryption key using the attack algorithm. Finally, this attack can cause revoked users to obtain sharing data and to disclose other secrets of legitimate members. Zhou et al. [14] presented a scheme of secure access control over encrypted data in cloud storage by invoking a role-based encryption technique. It is argued that the system can achieve effective user revocation that combines role-based access control policies with encryption to secure the storage of large data in the cloud. Unfortunately, inter-entity checks are not involved, the system easily suffers from attacks, for example, a collusion attack. Finally, this attack can lead to the disclosure of sensitive data files. Zou et al. [15] presented a convenient and flexible key management mechanism for trusted collaborative computing. By exploiting the access control polynomial, it is designed to provide effective access control for dynamic groups. Unfortunately, the secure way to share the permanent personal secret between the user and the server is not supported and the private key will be disclosed once the mobile secret p Permanent staff will have been obtained by the attackers. Nabeel et al. [16] proposed a content sharing scheme based on a privacy policy in public clouds. However, this system is not secure due to the low protection of the commitment in the identity token issuance phase.

II. RELATED WORK

Distributed computing is the conveyance of PC benefits over the Internet. Regardless of whether they understand it or not, many individuals utilize distributed computing administrations for their own needs. Here, keeping up information protection and character classification is a genuinely troublesome undertaking in multi-inhabitant information sharing. In this article, we propose a safe multi-proprietor information sharing plan by exploiting bunch marking and utilizing dynamic communicate encryption procedures that all individuals can share and information with different clients. What's more, here, the quantity of clients repudiated is free of the cost of capacity per head and encryption. In this article, the primary objective is to guarantee information security and show the adequacy of our plan in tests [13]. Distributed computing is a rising registering worldview in which IT foundation assets are given as Internet administrations. As promising as it might be, this worldview presents numerous new difficulties for information security and access control when clients outsource touchy information to share it on cloud servers that are not in an indistinguishable space from information proprietors. . To protect the privacy of touchy client information against untrusted servers, existing arrangements normally apply cryptographic techniques by uncovering the unscrambling keys just to approved clients. In any case, in doing as such, these arrangements definitely present an overwhelming registering overhead on the information proprietor for key circulation and information administration when fine granularity information get to control is wanted, and along these lines not develop well. The issue of at the same time acquiring the precision, adaptability, and classification of access control information is as yet uncertain [6]. In the distributed computing condition, putting away delicate information is a more troublesome errand. The cost of secrecy is high when we scramble whole delicate information. Encryption information does not function admirably in the cloud application. It has turned into the test to safeguard delicate information in the cloud. We in this manner dissect the information that must be scrambled and the others are most certainly not. And furthermore separate the information into various parts and put away in an alternate cloud condition. Each piece of the datasets contains the tokens. The capacity server distinguishes the information utilizing emblematic keys. This enables protection to save information assaults from assailants [5]. Security and protection are real worries in the reception of cloud advances for information stockpiling. One way to deal with moderate these worries is the utilization of two-layer encryption (TLE) which incorporates coarse granularity and fine granularity get to control encryption. Be that as it may, in this approach, information proprietors acquire high correspondence and figuring costs. To beat this issue, the information proprietor makes encryption identified with protection; while the Trusted Third Party (TTP) performs fine re-encryption on exclusive scrambled information that tackles this issue by utilizing limit based access control with TTP to guarantee that legitimate clients get to the outsourced information. In this article, we have proposed an encryption strategy at

TTP level to ensure the protection and respectability of outsourced information in the cloud condition [16].

III. Proposed Method

ASymmetric Key Management The first and oldest, based on key administration engineering of the 1970s, uses similar information encryption innovation to oversee keys and scramble information. In these frameworks, called "symmetric key" frameworks in light of the fact that a similar key is used to encode and decrypt data, the key manager produces another key for each message at the request of the sender. The key is stored in a database next to the list of collectors. At the time the beneficiary confirms, the key is retrieved from the database and the name of the collector is coordinated against the list of approved beneficiaries. In the case of everything looks, the descrambling key is sent to the receiver. Symmetric key systems have become the centerpiece of the inside just encryption and confirmation frameworks. Until now, Kerberos frameworks and Windows space controllers were based on symmetric key administration. The ability to quickly interpret key passwords and, to a large extent, the rapid execution of symmetric key encryption computations make these frameworks attractive for internal applications that do not need to incorporate any external clients into the encryption process. High storage costs - Many symmetric keyframes, but not all, require that a database containing the key for each message be available in the framework. While some advocates of symmetric keyframes will require that this database not be a significant obstruction, this key database must be reproduced, descended and supervised for the most part. Since this database contains basic security data (to be specific, the keys), these expenses are magnified. High Availability Requirements - Because the sender must request a key for each key manager message, the key supervisor is committed to each encryption operation. This implies that the key administrator must be exceptionally accessible and that the size of the key manager will limit the size of the entire information or information encryption framework. It also tends to increase the impact of capacity needs. B. Public Key Infrastructure (PKI) Key Management In the mid-1980s, an evolution of digital developments led to new and important types of encryption computations. These calculations, called "public keys" or "asymmetric", use an alternative key to scramble the information that they use to decrypt the information. The well-known Diffie-Hellman and RSA calculations are the best-known cases of open key calculations. While they are unable to encode substantial amounts of information, unbalanced calculations are ideally suited to key monitoring because they can quickly scramble smaller sized items. The basic thinking behind using open key calculations to monitor mass encryption keys is that a recipient produces two keys: an open key and a private key. To encode information, the sender creates a mass encryption key, scrambles the mass key with the recipient's open key, and sends the information next to the newly encrypted mass key. The recipient obtains the information, decrypts the mass key with his private key, and then uses that key to decode the data. On the surface, executives in the light of the key company's general base, or PKI, seem to understand the most pressing flaws of symmetric key frameworks: there is no requirement for a key database messages and the key server must not be reached for each message. Be that as it may, PKI has two notable obstacles to its ancestor, the symmetric key administration: 1) the creation of the private key at the beneficiary makes recovery of keys difficult to achieve and 2) the sender must find an open key for each beneficiary and confirms its legitimacy. Since the recipients create these keys themselves, the server will probably be unable to provide keys for all recipients. This failure to discover the keys for each recipient has made the administration of scrambled messaging frames, for example PGP and S / MIME impossible to encrypt. . This is the problem that authentications had to illuminate.

IV. CONCLUSION AND FUTURE WORK

In this article, we design a secure anti-collusion data sharing scheme for dynamic groups in the cloud. In our schema, users can securely obtain their private keys from the group manager and secure communication channels. In addition, our system is able to effectively support dynamic groups, when a new user joins the group or a user is revoked from the group, the private keys of other users do not need to be recalculated and put up to date. In addition, our system can perform a secure user revocation, revoked users can not get the original data files once they are revoked even if they conspire with the unreliable cloud. In this article, I can use an identity-based encryption algorithm, but in the future more new secure encryption techniques will be used and revoked users may be available but they will not be able to obtain the original data files.

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, Joseph AD, Katz R., Konwinski A., Lee G., Patterson D., Rabkin A., Stoica · I., and Zaharia M., of cloud computing, ‖ Common. ACM, vol. 53, no. 4, pp. 50-58, April 2010.

[2] S. Kamara and K. Lauter, "Cryptographic Cloud Storage", in Proc. Int. Conf. Financial Cryptography Data Security, January 2010, pp. 136-149.

[3] Mr. Kallahalla, E. Riedel, R. Swaminathan, Q. Wang, and K. Fu, -Plutus: Scalable Secure File Sharing on Unreliable Storage, ‖ in Proc. USENIX Conf. File Storage Technol., 2003, pp. 29-42.

[4] E. Goh, H. Shacham, N. Modadugu, and D. Boneh, -Sirius: Securing remote unsecured storage, ‖ in Proc. Netw. Distribute Syst. Security Symp., 2003, pp. 131-145.

[5] G. Ateniese, K. Fu, M. Green, and S. Hohenberger, -Improved proxy re-encryption schemes with applications to secure distributed storage, in Proc. Netw. Distribute Syst. Security Symp., 2005, pp. 29-43.

[6] S. Yu, Wang C., Ren K., and W. Lou, -Achieving secure, scalable and fine-grained data access control in cloud computing, ‖ in Proc. ACM Symp. Inf., Comput. Common. Security, 2010, pp. 282-292.

[7] V. Goyal, O. Pandey, A. Sahai, and B. Waters, - Attribute-based Encryption for Fine Granular Access Control of Encrypted Data, in Proc. ACM Conf. Comput. Common. Security, 2006, pp. 89-98.

[8] Lu R., X. Lin, X. Liang, and X. Shen, -Safe Source: The Essence of the Bread and Butter of Legal Computing in Cloud Computing, ‖ in Proc. ACM Symp. Inf., Comput. Common. Security, 2010, pp. 282-292.

[9] B. Waters, encryption based on the -Ciphertext-policy attribute: An expressive, efficient and surely secure realization, ‖ in Proc. Int. Conf. Theory of Practice Public Key Cryptography Conf. Public key cryptography, 2008, pp. 53-70.

[10] X. Liu, Y. Zhang, Wang B., and J. Yang, -Mona: Secure multi-proprietary data sharing for dynamic groups in the cloud, ‖ IEEE Trans. Distribute in parallel Syst., Vol. 24, no. 6, pp. 1182-1191, June 2013.

[11] D. Boneh, X. Boyen, and E. Goh, -Encryption based on hierarchical identity with a constant-size ciphertext, ‖ in Proc. Annu. Int. Conf. Theory Appl. Cryptographic Techn., 2005, pp. 440-456.

[12] C. Delerablee, P. Paillier, and D. Pointcheval, "Encryption of secure dynamic diffusion by integral collusion with Ci paradigms or decryption keys of constant size", in Proc. 1st Int. Conf. PairingBased Cryptography, 2007, pp. 39-59.

[13] Z. Zhu, Z. Jiang, and R. Jiang, -The attack on mona: secure multi-owner data sharing for dynamic groups in the cloud, ‖ in Proc. Int. Conf. Page 5841 Inf. Sci. Cloud Comput., December 7, 2013, pp. 185-189.

[14] L. Zhou, V. Varadharajan, and M. Hitchens, -Achieving access control based on secure role on encrypted data in cloud storage, ‖ IEEE Trans. Inf. Forensics Security, vol. 8, no. 12, pp. 1947-1960, dec. 2013.

[15] X. Zou, Y.-S. Dai, and E. Bertino, "A Practical and Flexible Key Management Mechanism for Trusted Collaborative Computing," in Proc. IEEE Conf. Comput. Commun., 2008, pp. 1211-1219.

[16] M. Nabeel, N. Shang, and E. Bertino, -Privacy preserving policy-based content sharing in public clouds, ‖ IEEE Trans. Know. Data Eng., Vol. 25, no. 11, pp. 2602-2614, Nov. 2013.

[17] D. Dolev and A. C. Yao, "On Public Key Protocol Security", IEEE Trans. Inf. Theory, vol. IT-29, no. 2, pp. 198-208, March 1983.

[18] B. Dan and F. Matt, "identity-based encryption from weil matching," in Proc. 21st Annu. Int. Cryptol. Conf. Adv. Cryptol., 2001, vol. 2139, pages 213-229.

[19] B. Den Boer, -Diffie-Hellman is as loud as discrete log for some prime numbers

International Journal of Research (IJR)

PEN2PRINT®

Certificate of Publication

is awarded to

NAVANEEHA POLEPALLY

Circuit Cipher text-Policy Attribute-Based Hybrid Encryption with Verifiable Delegation in Cloud Computing

Published in *International Journal of Research (IJR),* Vol-04, Issue-14
November 2017 ISSN: 2348-6848

International Refereed and Indexed Journal for Research Publication

With Impact Factor 5.60 UGC APPROVED journal Sr No. 44396

Index Copernicus Value (ICV) 100 & Indexed in Thomson Reuters

S.N. Sharma
Editor-in-Chief
International Journal of Research (IJR)
www.internationaljournalofresearch.com
Email: contacte@internationaljournalofresearch.com

ISSN 2348-6848
9 772348 684006 >

EduPub™

International Journal of Research (IJR)

PEN2PRINT®

PEN2PRINT®

Certificate of Publication

is awarded to

SATISH GUPTA BOYINA

Circuit Cipher text-Policy Attribute-Based Hybrid Encryption with Verifiable Delegation in Cloud Computing

Published in *International Journal of Research (IJR),* Vol-04, Issue-14
November 2017 ISSN: 2348-6848

International Refereed and Indexed Journal for Research Publication
With Impact Factor 5.60 UGC APPROVED journal Sr No. 44396
Index Copernicus Value (ICV) 100 & Indexed in Thomson Reuters

S.N. Sharma
Editor-in-Chief
International Journal of Research (IJR)
www.internationaljournalofresearch.com
Email: contacte@internationaljournalofresearch.com

ISSN 2348-6848
9 772348 684006 >

EduPub™

CIRCUIT CIPHERTEXT-POLICY ATTRIBUTE-BASED HYBRID ENCRYPTION WITH VERIFIABLE DELEGATION IN CLOUD COMPUTING

1. NAVANEEHA POLEPALLY, 2. SATISH GUPTA BOYINA

1.Pg Scholar, Department Of CSE, Annamacharya Institute Of Technology And Sciences,Piglipur, Batasingaram(V), Hayathnagar(M), Ranga Reddy(D),Hyderabad.

2. Asst.Professor and Head of the Department, Department Of CSE, Annamacharya Institute Of Technology And Sciences,Piglipur, Batasingaram(V), Hayathnagar(M), Ranga Reddy(D),Hyderabad

ABSTRACT: In the cloud, to give get to control and information security, information proprietors can utilize ascribe based encryption to scramble the put away information. In any case, to lessen the cost, clients with constrained processing power will probably assign the veil of the unscrambling errand to the cloud servers. The outcome demonstrates the encryption in light of properties with assignment. In any case, there are a few issues and inquiries with respect to the past related work. For instance, amid designation or production, servers in the cloud can speak to or supplant the appointed ciphertext and react to a false outcome with vindictive expectation. Notwithstanding cost reserve funds, the cloud server can likewise dupe qualified clients by disclosing to them that they are not dependable. Indeed, even access arrangements may not be adaptable amid encryption. Since the general circuit approach is utilized to get the most grounded type of access control, crossover encryption in view of the characteristics of the encryption strategy of the plan circuit has been encoded with irrefutable appointment. created. This framework is joined with an evident count and a Mac-then-Mac instrument, information protection, fine granular access control and the exactness of delegate PC comes about are ensured in the meantime. As this plan understands the security against chose assaults clear messages under the speculation of Diffie-Hellman Decisional kmultilinear. Furthermore, this plan accomplishes achievability and proficiency.

Watchwords: Encryption in view of encryption approach properties, circuits, obvious designation, multi-direct guide, half breed encryption.

I. INTRODUCTION

Cloud computing is innovation which uses advanced computational power as well as improved storage capabilities. Cloud computing is a long dreamed vision of computing utility, which enable the sharing of services over the internet. Cloud is a large group of interconnected computers, which is a major change in how we store information and run application. Cloud computing is a shared pool of configurable computing resources, on-demand network access and provisioned by the service provider. The advantage of cloud is cost savings. The prime disadvantage is security. The appearance of cloud computing transports a radical novelty to the organization of the data possessions within this calculating surroundings, the cloud servers can present different data services, such as isolated data storage and outsourced allocation calculation etc. For information cargo space, the servers amass a huge quantity of communal information, which might be accessed by certified users. For allocation calculation, the servers could be accustomed to hold and determine frequent data dealing to the user's burden. As applications shift to cloud computing proposals, verifying delegation process using cipher text-policy attribute-based encryption (CP-ABE) is used to guarantee the data privacy and the verifiability of allocation on untruthful cloud servers. Captivating health check data distribution as an example among the rising volumes of health check images and health check records, the medical care associations set a big amount of data in the cloud for dropping. To make such data sharing be achievable, attribute based encryption is used. There are two forms of attribute-based encryption. One is key-policy attribute-based encryption (KP-ABE) and the second is ciphertext-policy attribute-based encryption. In CP-ABE system, each ciphertext is contains an access structure, and each private key is labeled with a set of descriptive attributes. A user is able to decrypt a ciphertextif and only if the key's attribute set satisfies the access structure associated with a ciphertext. The cloud server provides another service which is delegation computing.The VD-CPABE schemeshows that the untrusted cloud will not be able to learn anything about the encrypted message and build the original ciphertext.

II. LITERATURE SURVEY Number Paper Name Author Name Proposed System Referred Point 1. Attribute-Based Access Control with EfficientRevocation in Data Outsourcing Systems JunbeomHur and Dong Kun Noh. In this paper,propose an access control mechanism using ciphertext-policy attribute-based encryption to enforce access control policies withefficient attribute and user revocation capability. In this Paper, we referred the solutionattribute-based encryption and selective group key distribution in each attribute group. 16. 0410178 18956 2. Privacy-preserving decentralized key-policy attribute-based Encryption J. Han, W. Susilo, Y. Mu, and J. Yan. In this paper, propose a privacy-preserving decentralized keypolicy ABE scheme where each authority can issue secretkeys to a user independently without knowing anything about his GID. In this Paper, we referred the solution the first decentralized ABE scheme with privacy-preserving based on standardcomplexity assumptions. 3. Securely outsourcing attribute-based encryption with checkability J. Li, X. Huang, J. Li, X. Chen, and Y. Xiang. This paper proposes anoutsourced ABE construction which providescheckability of the outsourced computation results in an efficient way. Extensive Security and performance analysis show that the proposed schemes are proven secure and practical. In this Paper, we referred the solution ABE .with verifiable delegation. Since the introduction of ABE, there have been advances in multiple directions. 4. A new paradigm of hybridencryption scheme K. Kurosawa and Y. Desmedt. In this paper, we show that a key encapsulation mechanism (KEM) does not have to be IND-CCA secure in the construction of hybrid encryption schemes, as was previously believed In this Paper, we have referred the solution to develop the KEM/DEM model for hybrid encryption. 5. A practical public key cryptosystemprovably secure against adaptive chosen ciphertext attack R. Cramer and V. Shoup. A new public key cryptosystem is proposed and analyzed. The scheme is quite practical, and is provably secure against adaptive chosenciphertext attack under standard intractability assumptions. In this paper, we have referred the solution to present and analyze a new public key cryptosystem that is provably secure against adaptive chosen ciphertext attack 16. 0410178 18957 6. Attribute-based encryptionwith verifiable outsourced decryption J. Lai, R. H. Deng, C. Guan, and J. Weng. In this Paper we proposed ABE system with outsourced decryption largely eliminates the decryption overhead of server. In such system, the proxy server such as cloud service provider is present which has a transformation key In this Paper , we referred the solution to the cloud servers canoffer various data services, such as outsourced delegation computation. 7. Ciphertext-policy attribute-based encryption: Anexpressive, efficient, and provably secure realization B. Waters. In this Paper, we proposed the solutions allow any encryptor to specify access control in terms of any access formula over the attributes in the system. In this Paper, we referred the solution to ensure the data confidentiality and the verifiability of delegation on dishonest cloud servers. 8. Decentralizing attributebasedencryption A. Lewko and B. Waters. In this Paper, We propose a MultiAuthority AttributeBased Encryption (ABE) system. In our system,any party can become an authority and there is no requirement for any global coordination other than the creation of an initial set of common reference parameters. In this Paper, we referred the solution to ABE authority by creating a public key and issuing private keys to different users that reflect their attributes. 9. How to delegateand verify in public: Verifiable computation from attributebasedencryption B. Parno, M. Raykova, and V. Vaikuntanathan. In this Paper, we Proposed the public delegation and public verifiability, which have important applications in many practical delegation scenarios In this Paper , we referred the solution the verifiability of delegation on dishonest cloud servers. 10. Outsourcing thedecryption of ABE Ciphertexts. M. Green, S. Hohenberger, and B. Waters. In this Paper, we propose a new paradigm for ABE that largely eliminates this overhead for users. In this Paper, we referred the solution the cloud servers can offer various data services and outsourced delegation computation.

III. EXISTING SYSTEM

In existing system,the attribute-based encryptiontechnique was used. But this scheme contains some problems and questions regarding to related works. Like during the delegation or releasethe cloud servers could misrepresent or replace the delegated cipher text and respond a fake result with malevolent intent. For the purpose of cost saving the cloud server may also fraud the eligible users by responding them that they are unworthy. Even, the access policies may not be flexible enough as well during the encryption. Disadvantage of Existing System:- No guarantee that the calculated result returned by the cloud is always correct. The cloud server may buildciphertext or fraud the eligible user that he even does not have permissions to•decryption. Loss the data security, confidentiality as well as access control.

IV. PROPOSED SYSTEM

The proposed framework, outline a circuit ciphertext-arrangement property based half and half encryption with obvious designation conspire. In this plan the circuits are utilized which express the most grounded type of access control strategy. The kmultilinear Decisional Diffie-Hellman supposition demonstrates the proposed plot is secure. On the opposite side, this plan can be valuable over the numbers. And also amid the appointment figuring, a client could approve whether the cloud server reacts a right changed ciphertext to help him/her decode the ciphertext quickly and accurately. Favorable position of Proposed System:- The nonexclusive KEM/DEM development for half breed encryption which can scramble messages of subjective length. Gives ensure for accuracy of the first ciphertext by utilizing a commitment.Achieves security, secrecy and additionally get to control The framework contains four modules,

1. Distributed storage Module

2. Information Owner Module

3. Information User Module

4. Specialist Module Cloud Storage

These distributed storage suppliers are in charge of keeping the information accessible and available, and the physical condition secured and running. Individuals and associations purchase or rent stockpiling limit from the suppliers to store end client, association, or application information. Information Owner: The information proprietor scrambles his message under access approach, at that point registers the supplement circuit, which yields the inverse piece of the yield of f, and encodes an arbitrary component R of a similar length to under the strategy Data User: The clients can outsource their perplexing access control arrangement choice and part procedure of decoding to the cloud. Which broadened encryption guarantees that the clients can acquire either the message M or the arbitrary component R, which stays away from the situation when the cloud server misleads the clients that they are not fulfilled to the entrance approach, notwithstanding, they meet the entrance strategy really. Expert: Authority produces private keys for the information proprietor and client.

V. CONCLUSION

Outline a circuit ciphertext-approach characteristic based cross breed encryption with provable portion strategy. The all inclusive circuits are useful to accomplish or clear the most grounded type of entrée oversee technique. Aggregate provable count and encode then-Mac framework with our ciphertext approach property based half and half encryption, we could appoint the provable fragmentary unscrambling worldview to the cloud server. The k-multilinear Decisional Diffie-Hellman suspicion demonstrates the plan is secure.On the opposite side, this plan can use over the whole numbers. The conclusion demonstrate that the technique is sensible in the distributed computing. Along these lines, can have the capacity to accomplish information security, the fine-grained entrée oversees and the evident distribution in cloud.

REFERENCES

[1]JunbeomHur and Dong Kun Noh," Attribute-Based Access Control with Efficient Revocation in Data Outsourcing Systems", VOL. 22, NO.7, JULY 2011 IEEE.

[2] J. Han, W. Susilo, Y. Mu, and J. Yan, "Protection saving decentralized key-approach property based Encryption," IEEE Trans. ParallelDistrib. Syst., vol. 23, no. 11, pp. 2150–2162, Nov. 2012.

[3] J. Li, X. Huang, J. Li, X. Chen, and Y. Xiang, "Safely outsourcing property based encryption with checkability," IEEE Trans. ParallelDistrib. Syst., vol. 25, no. 8, pp. 2201–2210, Aug. 2013.

[4] K. Kurosawa and Y. Desmedt, "another worldview of hybridencryption plot," in

Proc. 24th Int. Cryptol. Conf., 2004, pp. 426– 442.

[5] R. Cramer and V. Shoup, "A functional open key cryptosystemprovably secure against versatile picked ciphertext assault," inProc. eighteenth Int. Cryptol. Conf., 1998, pp. 13– 25.

[6] J. Lai, R. H. Deng, C. Guan, and J. Weng, "Property based encryptionwith unquestionable outsourced decoding," IEEE Trans. Inf. ForensicsSecur., vol. 8, no. 8, pp. 1343– 1354, Aug. 2013.

[7] B. Waters, "Ciphertext-strategy property based encryption: Anexpressive, productive, and provably secure acknowledgment," in Proc.14th Int. Conf. Practice Theory Public Key Cryptography. Conf. PublicKeyCryptography., 2011, pp. 53– 70.

[8] A. Lewko and B. Waters, "Decentralizing quality basedencryption," in Proc. 30th Annu. Int. Conf. Hypothesis Appl. Cryptograph.Techn., 2011, pp. 568–588.

[9] B. Parno, M. Raykova, and V. Vaikuntanathan, "How to delegateand check in broad daylight: Verifiable calculation from trait basedencryption," in Proc. ninth Int. Conf. Hypothesis Cryptograph., 2012,pp. 422– 439.

[10] M. Green, S. Hohenberger, and B. Waters, "Outsourcing thedecryption of ABE Ciphertexts," in Proc. USENIX Security Symp.,San Francisco, CA, USA, 2011, p. 34

IJR ® International Journal of Research (IJR)

PEN2PRINT®

PEN2PRINT®

Certificate of Publication

is awarded to

SHAIK NURJAHA

Nearest Keyword Set Search in Multidimensional Datasets

Published in *International Journal of Research (IJR),* Vol-04, Issue-14
November 2017 ISSN: 2348-6848

International Refereed and Indexed Journal for Research Publication

With Impact Factor 5.60 UGC APPROVED journal Sr No. 44396

Index Copernicus Value (ICV) 100 & Indexed in Thomson Reuters

S.N. Sharma
Editor-in-Chief
International Journal of Research (IJR)
www.Internationaljournalofresearch.com
Email: contacte@internationaljournalofresearch.com

ISSN 2348-6848
9 772348 684006 >

EduPub™

PEN2PRINT®

IJR ®

International Journal of Research (IJR)

PEN2PRINT®

Certificate of Publication

is awarded to

K. SANDHYA RANI

Nearest Keyword Set Search in Multidimensional Datasets

Published in *International Journal of Research (IJR),* Vol-04, Issue-14
November 2017 ISSN: 2348-6848

International Refereed and Indexed Journal for Research Publication

With Impact Factor 5.60 UGC APPROVED journal Sr No. 44396

Index Copernicus Value (ICV) 100 & Indexed in Thomson Reuters

S.N. Sharma
Editor-in-Chief
International Journal of Research (IJR)
www.internationaljournalofresearch.com
Email: contacte@internationaljournalofresearch.com

ISSN 2348-6848
9 772348 684006 >

EduPub™

NEAREST KEYWORD SET SEARCH IN MULTIDIMENSIONAL DATASETS

1. SHAIK NURJAHA, 2. K.SANDHYA RANI

1.Pg Scholar, Department Of CSE, Annamacharya Institute Of Technology And Sciences,Piglipur, Batasingaram(V), Hayathnagar(M), Ranga Reddy(D),Hyderabad.

2. Assoc.Professor and Head of the Department, Department Of CSE, Annamacharya Institute Of Technology And Sciences,Piglipur, Batasingaram(V), Hayathnagar(M), Ranga Reddy(D),Hyderabad

ABSTRACT:Keyword-based search in text-rich multi-dimensional datasets facilitates many novel applications and tools. In this paper, weconsider objects that are tagged with keywords and are embedded in a vector space. For these datasets, we study queries that ask forthe tightest groups of points satisfying a given set of keywords. We propose a novel method called ProMiSH (Projection and Multi ScaleHashing) that uses random projection and hash-based index structures, and achieves high scalability and speedup. We present anexact and an approximate version of the algorithm. Our experimental results on real and synthetic datasets show that ProMiSH has upto 60 times of speedup over state-of-the-art tree-based techniques.

1 INTRODUCTION

In today's digital world the amount of data which is developed is increasing day by day. There Is different multimedia in which data is saved. It's very difficult to search the large dataset for a given query as well to archive more accuracy on user query. In the same time query will search on dataset for exact keyword match and it will not find the nearest keyword for accuracy. Ex: Flickr.The amount of data which is developed is increasing day by day, thus it is very difficult to search large dataset for a given query as well to achieve more accuracy on user query.so we have implemented a method of efficient search in multidimensional dataset.This is associated with images as an input. Images are often characterized by a collection of relevant features, and are commonly represented as points in a multi dimensional feature space. For example, images are represented using colour feature vectors, and usually have descriptive text information (e.g., tags or keywords) associated with them. We consider multi dimensional datasets where each data point has a set of keywords. The presence of keywords in feature space allows for the development of new tools to query and explore these multi dimensional datasets Our main contributions are summarized as follows.

(1) We propose a novel multi scale index for exact and Approximate NKS query processing.

(2) We develop efficient search algorithms that work with

the multi scale indexes for fast query processing.(3) We conduct extensive experimental studies to demonstrate the performance of the proposed techniques.

1. Filename:It is based on image filename.

2. CBIR (Content based image search): Content based image retrieval (CBIR), also known as query by image content (QBIC) and content based visual information retrieval (CBVIR) is the application of computer vision techniques to the image retrieval problem, that is, the problem of searching for digital images in large databases. Content based image retrieval is opposed to traditional concept based approaches (see Concept based image indexing).

3. TBIR (Text based image search): Concept based image indexing, also variably named as “description based” or“text based” image in dexing/retrieval, refers to ret rieval from text based indexing of images that may employ keywords, subject headings, captions, or natural language text. It is opposed to Content based image retrieval. Indexing is a technique used in CBIR.

2. LITERATURE SURVEY

We study nearest keyword set (referred to as NKS) queries on text rich multi dimensional datasets. An NKS query is a set of user provided keywords, and the result of the query may include k sets of data points each of which contains all the query keywords and forms one of the top-k tightest clusters in the multi dimensional space. Illustrates an NKS query over a set of two dimensional data points. Each point is tagged with a set of keywords. For a query the set of points contains all thequery keywords and forms the tightest cluster compared with any other set of points covering all the query keywords. Therefore, the set is the International Research result for the query Q.NKS queries are useful for many applications, such as photo sharing in social networks, graph pattern search, geolocation search in GIS systemsand so on.We present an exact and an approximate version of the algorithm. Our experimental results on real and synthetic datasets show that the method has more speedup over state of the art tree based techniques.Other related queries include aggregate nearest keyword search in spatial databases, top k preferential query, top k sites in a spatial data based on their influence on feature points, and optimal location queries. Our work is different from these techniques. First, existing works mainly focus on the type of queries where the coordinates of query points are known. Even though it ispossible to make their cost functions same to the cost function in NKS queries, suchtuning does not change their techniques. The proposed techniques use location information as an integral part to perform a best first search on the IR Tree, and query coordinates play a fundamental role in almost every step of the algorithms to prune the search space. Moreover, these techniques do not provide concrete guidelines on how to enable efficient processing for the type of queries where query coordinates are missing. Second, in multi dimensional spaces, it is difficult for users to provide meaningful coordinates, and our work deals with another type of queries where users can only provide keywords as input. With out query coordinates, it is difficult to adapt existing techniques to our problem.Finding nearest neighbors in large multi dimensional data has always been one of the research interests in data mining field. In this paper, we present our continuous research on similarity search problems. Previous work on exploring the meaning of K nearest neighbors from a new perspective in Pan KNN. It redefines the distances between data points and a given query point Q, efficiently and effectively selecting data points which are closest to Q. It can be applied in various data mining fields. A large amount of real data sets have irrelevant or obstacle information which greatly affects the effectiveness and efficiency of finding nearest neighbors for a given query data point. In this paper, we present our approach to solving the similarity search problem in the presence of obstacles. We apply the concept of obstacle points and process the similarity search problems in a different way. This approach can assist to improve the performance of existing data analysis approaches.

The similarity between two data points used to be based on a similarity function such as Euclidean distance which aggregates the difference between each dimension of thetwo data points in traditional nearest neighbor problems.In those applications, the nearest neighbor problems are solved based on the distance between the data point and the query point over a fixed set of dimensions (features). However, such approaches only focus on full similarities, i.e., the similarity in full data space of the data set. Also early methods suffer from the "curse of dimensionality". In a high dimensional space the data are usually sparse, and widely used distance metric such as Euclidean distance may not work well as dimensionalitygoes higher. Recent research [8] shows that in high dimensions nearest neighbor queries become unstable: the difference of the distances of farthest and nearest points to some query point does not increase as fast as the minimum of the two, thus the distance between two data points in high dimensionality is less meaningful. Some approaches are proposed targeting partial similarities. However, they have limitations such as the requirement of the fixed subset of dimensions, or fixed number of dimensions as the input parameter(s) for the algorithms.

3 INDEX STRUCTURE FOR EXACT PROMISH

We start with the index for exact ProMiSH (ProMiSH-E). This index consists of two main components. Inverted Index Ikp. The first component is an inverted index referred to as Ikp. In Ikp, we treat keywords as keys, and each keyword points to a set of data points that are associated with the keyword. Let D be a set of data points and V be a dictionary that contains all the keywords appearing in D. We build Ikp for D as follows. (1) For each v 2 V, we create a key entry in Ikp, and this key entry points to a set of data points Dv ¼ fo 2 Dj v 2 sðoÞg (i.e., a set includes all data points in D that contain keyword v). (2) We repeat (1) until all the keywords in V are processed. In Fig. 2, an example for Ikp is shown in the dashed rectangle at the bottom. Hashtable-Inverted Index Pairs HI. The second component consists of multiple hashtables and inverted indexes referred to as HI. HI is controlled by three parameters: (1) (Index level) L, (2) (Number of random unit vectors) m, and (3) (hashtable size) B. All the three parameters are non-negative integers. Next, we describe how these three parameters control the construction of HI. In general, HI contains L hashtable-inverted index pairs,characterized by fðHðsÞ; IðsÞ khbÞ j s 2 f0; 1; 2; . . . ; L _ 1gg,where HðsÞ and IðsÞ khb are the s-th hashtable and inverted index, respectively.

Algorithm . SearchInSubset

In: F0: subset of points; Q: query keywords; q: query size

In: PQ: priority queue of top-k results

```
rk PQ½k_:r /* kth smallest diameter */
SL ½ðv; ½ _Þ_: list of lists to store groups per querykeyword
for all v 2 Q do
    SL½v_ f8o 2 F0 : o is tagged with vg /* form groups */
end for
/* Pairwise inner joins of the groups*/
AL: adjacency list to store distances between points
M 0: adjacency list to store count of pairs betweengroups
for all ðvi; vjÞ 2 Q such that i  q; j  q; i < j do
    for all o 2 SL½vi_ do
        for all o0 2 SL½vj_ do
            if jjo _ o0jj2  rk then
                AL½o; o0_ jjo _ o0jj2
                M½vi; vj_ M½vi; vj_ þ 1
            end if
        end for
    end for
end for
/* Order groups by a greedy approach */
curOrder ½ _
while Q 6¼ ; do
    ðvi; vjÞ removeSmallestEdge(M)
    if vi 62 curOrder then
        curOrder.append(vi); Q Q n vi
    end if
    if vj 62 curOrder then
        curOrder.append(vj); Q Q n vj
    end if
end while
sort(SL, curOrder) /* order groups */
```

31: findCandidates(q, AL, PQ, Idx, SL, curSet, curSetr, rk)

4.PROPOSED SYSTEM

In this paper, we consider multi-dimensional datasets where each data point has a set of keywords. The presence of keywords in feature space allows for the development of new tools to query and explore these multi-dimensional datasets.In this paper, we study nearest keyword set (referred to asNKS) queries on text-rich multi-dimensional datasets. AnNKS query is a set of user-provided keywords, and theresult of the query may include k sets of data points each ofwhich contains all the query keywords and forms one of thetop-k tightest cluster in the multi-dimensional space.we propose ProMiSH (short for Projection and Multi-Scale Hashing) to enable fast processing for NKS queries. In particular, we develop an exact ProMiSH (referred to as ProMiSH-E) that always retrieves the optimal top-k results, and an approximate ProMiSH (referred to as ProMiSH-A) that is more efficient in terms of time and space, and is able to obtain near-optimal results in practice. ProMiSH-E uses a set of hashtables and inverted indexes to perform a localized search.

5 CONCLUSIONS

In this paper, we proposed solutions to the problem of top-k nearest keyword set search in multi-dimensional datasets.We proposed a novel index called ProMiSH based on random projections and hashing. Based on this index, we developed ProMiSH-E that finds an optimal subset of points and ProMiSH-A that searches near-optimal results with better efficiency. Our empirical results show that ProMiSH is faster than state-of-the-art tree-based techniques, with multiple orders of magnitude performance improvement. Moreover, our techniques scale well with both real and synthetic datasets. Ranking functions. In the future, we plan to explore other scoring schemes for ranking the result sets. In one scheme, we may assign weights to the keywords of a point by using techniques like tf-idf. Then, each group of points can be scored based on distance between points and weights of keywords. Furthermore, the criteria of a result containing all the keywords can be relaxed to generate results having only a subset of the query keywords

REFERENCES

[1] W. Li and C. X. Chen, "Efficient data modeling and querying system for multi-dimensional spatial data," in Proc. 16th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst., 2008, pp. 58:1–58:4.

[2] D. Zhang, B. C. Ooi, and A. K. H. Tung, "Locating mapped resources in web 2.0," in Proc. IEEE 26th Int. Conf. Data Eng., 2010, pp. 521–532.

[3] V. Singh, S. Venkatesha, and A. K. Singh, "Geo-clustering of images with missing geotags," in Proc. IEEE Int. Conf. Granular Comput., 2010, pp. 420–425.

[4] V. Singh, A. Bhattacharya, and A. K. Singh, "Querying spatial patterns," in Proc. 13th Int. Conf. Extending Database Technol.: Adv. Database Technol., 2010, pp. 418–429.

[5] J. Bourgain, "On lipschitz embedding of finite metric spaces in Hilbert space," Israel J. Math., vol. 52, pp. 46–52, 1985.

[6] H. He and A. K. Singh, "GraphRank: Statistical modeling and mining of significant subgraphs in the feature space," in Proc. 6th Int. Conf. Data Mining, 2006, pp. 885–890.

[7] X. Cao, G. Cong, C. S. Jensen, and B. C. Ooi, "Collective spatial keyword querying," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2011, pp. 373–384.

[8] C. Long, R. C.-W. Wong, K. Wang, and A. W.-C. Fu, "Collective spatial keyword queries: A distance owner-driven approach," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2013, pp. 689–700.

[9] D. Zhang, Y. M. Chee, A. Mondal, A. K. H. Tung, and M. Kitsuregawa, "Keyword search in spatial databases: Towards searching by document," in Proc. IEEE 25th Int. Conf. Data Eng., 2009, pp. 688–699.

[10] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Localitysensitive hashing scheme based on p-stable distributions," in Proc. 20th Annu. Symp. Comput. Geometry, 2004, pp. 253–262.

[11] Y. Zhou, X. Xie, C. Wang, Y. Gong, and W.-Y. Ma, "Hybrid index structures for location-based web search," in Proc. 14th ACM Int. Conf. Inf. Knowl. Manage., 2005, pp. 155–162.

[12] R. Hariharan, B. Hore, C. Li, and S. Mehrotra, "Processing spatialkeyword (SK) queries in geographic information retrieval (GIR) systems," in Proc. 19th Int. Conf. Sci. Statistical Database Manage., 2007, p. 16.

[13] S. Vaid, C. B. Jones, H. Joho, and M. Sanderson, "Spatio-textual indexing for geographical search on the web," in Proc. 9th Int. Conf. Adv. Spatial Temporal Databases, 2005, pp. 218–235.

[14] A. Khodaei, C. Shahabi, and C. Li, "Hybrid indexing and seamless ranking of spatial and textual features of web documents," in Proc. 21st Int. Conf. Database Expert Syst. Appl., 2010, pp. 450–466.

[15] A. Guttman, "R-trees: A dynamic index structure for spatial searching," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 1984, pp. 47–57.

[16] I. De Felipe, V. Hristidis, and N. Rishe, "Keyword search on spatial databases," in Proc. IEEE 24th Int. Conf. Data Eng., 2008, pp. 656–665.

[17] G. Cong, C. S. Jensen, and D. Wu, "Efficient retrieval of the top-k most relevant spatial web objects," Proc. VLDB Endowment, vol. 2, pp. 337–348, 2009.

A Survey Paper on Data Lineage in Malicious Environments

1. MEHABUNNISA

2. MS.K.NAGALATHA

1.Pg Scholar, Department Of ECE, Annamacharya Institute Of Technology And Sciences,Piglipur, Batasingaram(V), Hayathnagar(M), Ranga Reddy(D),Hyderabad.

2. Asst.Professor and Head of the Department, Department Of ECE, Annamacharya Institute Of Technology And Sciences,Piglipur, Batasingaram(V), Hayathnagar(M), Ranga Reddy(D),Hyderabad

ABSTRACT:

Intentional or unintentional leakage of confidential data is undoubtedly one of the most severe security threats thatorganizations face in the digital era. The threat now extends to our personal lives: a plethora of personal information is available tosocial networks and smartphone providers and is indirectly transferred to untrustworthy third party and fourth party applications. In thiswork, we present a generic data lineage framework LIME for data flow across multiple entities that take two characteristic, principal roles(i.e., owner and consumer). We define the exact security guarantees required by such a data lineage mechanism toward identificationof a guilty entity, and identify the simplifying non-repudiation and honesty assumptions. We then develop and analyze a novelaccountable data transfer protocol between two entities within a malicious environment by building upon oblivious transfer, robustwatermarking, and signature primitives. Finally, we perform an experimental evaluation to demonstrate the practicality of our protocoland apply our framework to the important data leakage scenarios of data outsourcing and social networks. In general, we consider LIME, our lineage framework for data transfer, to be an key step towards achieving accountability by design.

1 INTRODUCTION

IN the digital era, information leakage through unintentionalexposures, or intentional sabotage by disgruntledemployees and malicious external entities, present one ofthe most serious threats to organizations. According to aninteresting chronology of data breaches maintained by thePrivacy Rights Clearinghouse (PRC), in the United Statesalone, 868;045;823 records have been breached from 4;355data breaches made public since 2005 [1]. It is not hard tobelieve that this is just the tip of the iceberg, as most casesof information leakage go unreported due to fear of loss ofcustomer confidence or regulatory penalties: it costs companieson average $214 per compromised record [2]. Largeamounts of digital data can be copied at almost no cost andcan be spread through the internet in very short time. Additionally,the risk of getting caught for data leakage is verylow, as there are currently almost no accountability mechanisms.For these reasons, the problem of data leakage hasreached a new dimension nowadays.Not only companies are affected by data leakage, it isalso a concern to individuals. The rise of social networksand smartphones has made the situation worse. In theseenvironments, individuals disclose their personal informationto various service providers, commonly known as thirdparty applications, in return for some possibly free services.In the absence of proper regulations and accountabilitymechanisms, many of these applications share individuals'identifying information with dozens of advertising andInternet tracking companies.Even with access control mechanisms, where access tosensitive data is limited, a malicious authorized user canpublish sensitive data as soon as he receives it. Primitiveslike encryption offer protection only as long as the informationof interest is encrypted, but once the recipient decryptsa message, nothing can prevent him from publishing thedecrypted content. Thus it seems impossible to prevent dataleakage proactively.Privacy, consumer rights, and advocacy organizationssuch as PRC [3] and EPIC [4] try to address the problem ofinformation leakages through policies and awareness. However,as seen in the following scenarios the effectiveness ofpolicies is questionable as long as it is not possible to provablyassociate the guilty parties to the leakages.

2 RELATED WORK

A preliminary shorter version of this paper appeared at the STM workshop . This version constitutes a significantextension by including the following contributions:We give a more detailed description of our model, a formalspecification of the used primitives, an analysis ofthe introduced protocol, a discussion of implementationresults, an application of our framework to examplescenarios, a discussion of additional features and anextended discussion of related work.Clustering analysis is veryuseful to estimate the inter-entity similarity. One good example

of clustering based reranking algorithms is the InformationBottle based scheme developed by Hsu et al.[9]. In thismethod, the images in the initial results are primarily groupedautomatically into several clusters. Then the re-ranked resultlist is created first by ordering the clusters according tothe cluster conditional probability and next by ordering thesamples within a cluster based on their cluster membership value. In a fast and accurate scheme is proposed forgrouping Web image search results into semantic clusters. Itis obvious that the clustering based reranking methods canwork well when the initial search results contain many nearduplicate media documents. However, for queries that returnhighly diverse results or without clear visual patterns, theperformance is not guaranteed.

3 THE LIME FRAMEWORK

As we want to address a general case of data leakage in datatransfer settings, we propose the simplifying model LIME(Lineage in the malicious environment). With LIME weassign a clearly defined role to each involved party anddefine the inter-relationships between these roles. Thisallows us to define the exact properties that our transferprotocol has to fulfill in order to allow a provable identificationof the guilty party in case of data leakage.

3.1 Model

As LIME is a general model and should be applicable to allcases, we abstract the data type and call every data item document.There are three different roles that can be assigned tothe involved parties in LIME: data owner, data consumer andauditor. The data owner is responsible for the managementof documents and the consumer receives documents andcan carry out some task using them. The auditor is notinvolved in the transfer of documents, he is only invokedwhen a leakage occurs and then performs all steps that arenecessary to identify the leaker. All of the mentioned rolescan have multiple instantiations when our model is appliedto a concrete setting. We refer to a concrete instantiation ofour model as scenario.In typical scenarios the owner transfers documents toconsumers. However, it is also possible that consumers passon documents to other consumers or that owners exchangedocuments with each other.

In the outsourcing scenario [6]the employees and their employer are owners, while theoutsourcing companies are untrusted consumers.In the following we show relations between the differententities and introduce optional trust assumptions. We onlyuse these trust assumptions because we find that they arerealistic in a real world scenario and because it allows us tohave a more efficient data transfer in our framework. At theend of this section we explain how our framework can beapplied without any trust assumptions.When documents are transferred from one owner toanother one, we can assume that the transfer is governed bya non-repudiation assumption. This means that the sendingowner trusts the receiving owner to take responsibility ifhe should leak the document. As we consider consumersas untrusted participants in our model, a transfer involvinga consumer cannot be based on a non-repudiation assumption.Therefore, whenever a document is transferred to aconsumer, the sender embeds information that uniquelyidentifies the recipient. We call this fingerprinting. If the consumerleaks this document, it is possible to identify himwith the help of the embedded information.As presented, LIME relies on a technique for embeddingidentifiers into documents, as this provides an instrumentto identify consumers that are responsible for data leakage.We require that the embedding does not not affect the utilityof the document. Furthermore, it should not be possiblefor a malicious consumer to remove the embedded informationwithout rendering the document useless. A techniquethat can offer these properties is robust watermarking. Wegive a definition of watermarking and a detailed descriptionof the desired..

4 ACCOUNTABLE DATA TRANSFER

In this section we specify how one party transfers a documentto another one, what information is embedded andwhich steps the auditor performs to find the guilty party incase of data leakage. We assume a public key infrastructureto be present, i.e., both parties know each others signatureverification key.

4.1 Trusted Sender

In the case of a trusted sender it is sufficient for the sender toembed identifying information, so that the guilty party canbe found. As the sender is trusted, there is no need for furthersecurity mechanisms. we present a transferprotocol that fulfills the properties of correctness and nodenial as. As the sender is trusted tobe honest, we do not need the no framing property.The sender, who is in possession of some document D,creates a watermarking key k, embeds a triple consisting of the two parties' identifiers and a timestampt into D to create Dw ¼WðD; s; kÞ. He then sends Dwto the recipient, who will be held accountable for thisversion of the document. As the sender also knows Dw, thisvery simple protocol is only applicable if the sender iscompletely trusted; otherwise the sender could publish Dwand blame the recipient.

4.2 Untrusted Sender

In the case of an untrusted sender we have to take additionalactions to prevent the sender from cheating, i.e., wehave to fulfill the no framing property. To achieve this property,the sender divides the original document into n partsand for each part he creates two differently watermarkedversions. He then transfers one of each of these two versionsto the recipient via OT2

1 . The recipient is held accountableonly for the document with the parts that he received, butthe sender does not know which versions that are. Theprobability for the sender to cheat is therefore 12n. We showthe protocol and provide an analysis of the protocolproperties.First, the sender generates two watermarking keys k1 andk2. It is in his own interest that these keys are fresh and distinct.The identifying information that the sender embedsinto the documentD is a signed statement s ¼ ½CS; CR; t_skCRcontaining the sender's and recipient's identifiers and atimestamp t, so that every valid watermark is authorized bythe recipient. The sender computes the watermarked documentsplits the document D0 into n partsand creates two different versions

4.3 Data Lineage Generation

The auditor is the entity that is used to find the guilty partyin case of a leakage. He is invoked by the owner of the documentand is provided with the leaked document. In order toProtocol for trusted senders: The sender watermarks the originaldocument with a signed statment containing the participants' identifiersand a timestamp, and sends the watermarked document to the recipient. find the guilty party, the auditor proceeds in the followingway:

1) The auditor initially takes the owner as the currentsuspect.

2) The auditor appends the current suspect to thelineage.

3) The auditor sends the leaked document to the currentsuspect and asks him to provide the detectionkeys k1 and k2 for the watermarks in this documentas well as the watermark s. If a non-blind watermarkingscheme is used, the auditor additionallyrequests the unmarked version of the document.

4) If, with key k1, s cannot be detected, the auditor continueswith 9.

5) If the current suspect is trusted, the auditor checksthat s is of the form where CS is the identifierof the current suspect, takes CR as current suspectand continues with 2.

6) The auditor verifies that s is of the form ½CS;CR; t_skCRwhere CS is the identifier of the currentsuspect. He also verifies the validity of the signature.

7) The auditor splits the document into n parts and foreach part he tries to detect 0 and 1 with key k2. Ifnone of these or both of these are detectable, he continueswith 9. Otherwise he sets b0i as the detected bitfor the ith part. He sets b0 ¼ b01 . . . b0n.

8) The auditor asks CR to prove his choice of b ¼ b1 _ _ _ bn for the given timestamp t by presenting the. If CR is not able to give a correctproof (i.e., mi;bi is of the wrong form or the signatureis invalid) or if b ¼ b0, then the auditor takes CR ascurrent suspect and continues with 2.

9) The auditor outputs the lineage. The last entry isresponsible for the leakage.

CONCLUSION AND FUTURE DIRECTIONS

We present LIME, a model for accountable data transferacross multiple entities. We define participating parties,their inter-relationships and give a concrete instantiation fora data transfer protocol using a novel combination of obliviousrtransfer, robust watermarking and digital signatures.We prove its correctness and show that it is realizable bygiving microbenchmarking results. By presenting a generalapplicable framework, we introduce accountability as earlyas in the design phase of a data transfer infrastructure.Although LIME does not actively prevent data leakage, itintroduces reactive accountability. Thus, it will deter maliciousparties from leaking private documents and willencourage honest (but careless) parties to provide therequired protection for sensitive data. LIME is flexible as wedifferentiate between trusted senders (usually owners) anduntrusted senders (usually consumers). In the case of thetrusted sender, a very simple protocol with little overheadis possible. The untrusted sender requires a more complicatedprotocol, but the results are not based on trustassumptions and therefore they should be able to convincea neutral entity (e.g., a judge).Our work also motivates further research on dataleakage detection techniques for various document typesand scenarios. For example, it will be an interestingfuture research direction to design a verifiable lineageprotocol for derived data.

.REFERENCES

[1] Chronology of data breaches [Online]. Available: http://www.privacyrights.org/data-breach, 2014.

[2] Data breach cost [Online]. Available: http://www.symantec.com/about/news/release/article.jsp?prid=20110308_01, 2011.

[3] Privacy rights clearinghouse [Online]. Available: http://www.privacyrights.org, 2014.

[4] (1994). Electronic privacy information center (EPIC) [Online].Available: http://epic.org, 1994.

[5] Facebook in privacy breach [Online]. Available: http://online.wsj.com/article/SB10001424052702304772804575558484075236968.html, 2010.

[6] Offshore outsourcing [Online]. Available: http://www.computerworld.com/s/article/109938/Offshore_outsourcing_cited_in_Florida_data_leak, 2006.

[7] A. Mascher-Kampfer, H. St€ogner, and A. Uhl, “Multiple re-watermarkingscenarios,” in Proc. 13th Int. Conf. Syst., Signals, ImageProcess., 2006, pp. 53–56.

[8] P. Papadimitriou and H. Garcia-Molina, “Data leakage detection,”IEEE Trans. Knowl. Data Eng., vol. 23, no. 1, pp. 51–63, Jan. 2011.

[9] Pairing-based cryptography library (PBC) [Online]. Available:http://crypto.stanford.edu/pbc, 2014.

[10] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, “Secure spreadspectrum watermarking for multimedia,” IEEE Trans. ImageProcess., vol. 6, no. 12, pp. 1673–1687, Dec. 1997.

[11] B. Pfitzmann and M. Waidner, “Asymmetric fingerprintingfor larger collusions,” in Proc. 4th ACM Conf.

Comput. Commun.Security, 1997, pp. 151–160.

[12] S. Goldwasser, S. Micali, and R. L. Rivest, "A digital signaturescheme secure against adaptive chosen-message attacks," SIAMJ. Comput., vol. 17, no. 2, pp. 281–308, 1988.

[13] A. Adelsbach, S. Katzenbeisser, and A.-R. Sadeghi, "A computationalmodel for watermark robustness," in Proc. 8th Int. Conf. Inf.Hiding, 2007, pp. 145–160.

[14] J. Kilian, F. T. Leighton, L. R. Matheson, T. G. Shamoon, R. E.Tarjan, and F. Zane, "Resistance of digital watermarks tocollusive attacks," in Proc. IEEE Int. Symp. Inf. Theory, 1998,pp. 271–271.

[15] M. Naor and B. Pinkas, "Efficient oblivious transfer protocols," inProc. 12th Annu. ACM-SIAM Symp. Discrete Algorithms, 2001,pp. 448–457.

[16] GNU multiple precision arithmetic library (GMP) [Online]. Available:http://gmplib.org/, 2014.

[17] D. Boneh, B. Lynn, and H. Shacham, "Short signatures fromthe Weil pairing," in Proc. 7th Int. Conf. Theory Appl. Cryptol. Inf.Security: Adv. Cryptol., 2001, pp. 514–532.

[18] W. Dai. Crypto++ Library [Online]. Available: http://cryptopp.com, 2013.

[19] P. Meerwald. Watermarking toolbox [Online]. Available: http://www.cosy.sbg.ac.at/ pmeerw/Watermarking/source, 2010.

[20] Y. Ishai, J. Kilian, K. Nissim, and E. Petrank, "Extending oblivioustransfers efficiently," in Proc. 23rd Annu. Int. Cryptol. Conf. Adv.Cryptol., 2003, pp. 145–161.

Spoofer Location Detection Using Passive Ip Trace back

1. PALDE SUDHA JYOTHI 2. ARAVA NAGASRI

1.Pg Scholar, Department Of ECE, Annamacharya Institute Of Technology And Sciences,Piglipur, Batasingaram(V), Hayathnagar(M), Ranga Reddy(D),Hyderabad.

2. Asst.Professor and Head of the Department, Department Of ECE, Annamacharya Institute Of Technology And Sciences,Piglipur, Batasingaram(V), Hayathnagar(M), Ranga Reddy(D),Hyderabad

ABSTRACT

It is long known attackers may utilize fashioned source IP location to cover their real areas. To capture the spoofers, various IP traceback mechanisms have been proposed. However, due to the challenges regarding deployment services, there has been not any widely adopted IP traceback solution, at least at the Internet level. As a result, the mist on the locations of spoofers has never been dissolute till now. This paper proposes passive IP traceback (PIT) that bypasses the deployment difficulties of IP traceback techniques and comes up with a solution to the problem. PIT inve stigates Internet Control Message Protocol (ICMP) error messages (named path backscatter) triggered by spoofing traffic, and tracks the spoofers based on public available information such as topology. Along these lines, PIT can discover the spoofers with no arrangement necessity. This paper represents the reasons, accumulation, and the factual results on way backscatter, exhibits the procedures and adequacy of PIT, and demonstrates the caught areas of spoofers through applying PIT on the way backscatter information set. These results can help further reveal IP spoofing, which has been studied for long but never well understood. As because of some limitations PIT cannot work in all the spoofing attacks, it may be a helpful mechanism of tracing a spoofers before an Internet-level traceback system has been deployed in real. Keywords:- Computer network management, computer network security, denial of service (DoS), IP traceback.

I. INTRODUCTION

IP spoofing, which means attackers launching attacks with forged source IP addresses, has been recognized as a serious security problem on the Internet for long. By using addresses that are assigned to others or not assigned at all, attackers can avoid exposing their real locations thus protecting them from being traced, or enhance the effect of attacking, or launch reflection based attacks. A number of scandalous attacks rely on IP spoofing, including SYN flooding, SMURF, DNS amplification etc. A Domain Name System (DNS) amplification attack which severely degraded the service of a Top Level Domain (TLD) name server is reported in. Though there has been a popular conventional wisdom that DoS attacks [1] are launched from botnets and spoofing is no longer critical, the report of ARBOR on NANOG 50th meeting shows spoofing is still significant in observed DoS attacks. Indeed, based on the captured backscatter messages from UCSD Network Telescopes [2], spoofing activities are still frequently observed. To capture the origins of IP spoofing traffic is of great importance. As long as the actual and real locations of spoofers are not disclosed, they cannot be deterred, stopped and prevented from launching further attacks. Even just approaching the spoofers, for example, determining the ASes or networks they reside in, attackers can be located and traced in a smaller area, and filters can be placed and arranged closer to the attacker before attacking traffic get aggregated. The last but not the least, identifying the origins of spoofing traffic can help build a reputation system for ASes, which would be helpful to push the corresponding ISPs to verify IP source address [3]. This is the first article known which deeply investigates path backscatter messages. These messages are important and valuable to help understand and analyze the spoofing activities. Backscatter messages, which are produced and generated by the targets of spoofing messages, to study Denial of Services (DoS) [4] [5], path backscatter messages, which are sent by intermediate devices during the information exchange and transfer rather than the targets, have not been used in traceback. A practical and effective IP traceback solution based on path backscatter messages, i.e., PIT, is proposed. PIT bypasses the deployment difficulties of existing IP traceback RESEARCH ARTICLE OPEN ACCESS mechanisms [6] and actually is already in force. Though given the limitation that path backscatter messages are not generated with stable possibility, PIT cannot work in all the attacks, but it does work in a number of spoofing activities. At least it may be the most useful traceback mechanism before an AS-level traceback system has been deployed in real. Through applying PIT on the path backscatter dataset, a number of locations of spoofers are captured and presented. Though this is not a complete list, it is the first known list disclosing the locations of spoofers.

II. LITERATURE SURVEY

A. Efficient Packet Marking for Large-Scale

IP Traceback Author proposed a new approach to IP traceback based on the probabilistic packet marking paradigm [7]. Our approach, which we call randomize-and-link, uses large checksum cords to "link" message fragments in a way that is highly scalable, for the checksums serve both as associative addresses and data integrity verifiers. The main advantage of these checksum cords is that they spread the addresses of possible router messages across a spectrum that is too large for the attacker to easily create messages that collide with legitimate messages. Our methods therefore scale to attack trees containing hundreds of routers and do not require that a victim know the topology of the attack tree a priori. In addition, by utilizing authenticated dictionaries in a novel way, our methods do not require routers sign any setup messages individually.

B. Practical Network Support for IP Traceback This paper [8] describes a technique for tracing anonymous packet flooding attacks in the Internet back towards their source. This work is motivated by the increased frequency and sophistication of denial-of-service attacks and by the difficulty in tracing packets with incorrect, or ”spoofed”, source addresses. In this paper we describe a general purpose traceback mechanism based on probabilistic packet marking in the network. Our approach allows a victim to identify the network path(s) traversed by attack traffic without requiring interactive operational support from Internet Service Providers (ISPs) [3]. Moreover, this traceback can be performed ”post-mortem” after an attack has completed. We present an implementation of this technology that is incrementally deployable, (mostly) backwards compatible and can be efficiently implemented using conventional technology.

C. FIT: Fast Internet Traceback [9] E-crime is on the rise. The costs of the damages are often on the order of several billion of dollars. Traceback mechanisms are a critical part of the defense against IP spoofing and DoS attacks. Current traceback mechanisms are inadequate to address the traceback problem Problems with the current traceback mechanisms: • victims have to gather thousands of packets to reconstruct a single attack path • they do not scale to large scale attacks • they do not support incremental deployment General properties of FIT: • IncDep • RtrChg • FewPkt • Scale • Local D. ICMP Traceback with Cumulative Path, An Effcient Solution for IP Traceback DoS/DDoS attacks constitute one of the major classes of security threats in the Internet today. The attackers usually use IP spoofing to conceal their real location. The current Internet protocols and infrastructure do not provide intrinsic support to traceback the real attack sources. The objective of IP Traceback is to determine the real attack sources, as well as the full path taken by the attack packets. Different traceback methods have been proposed, such as IP logging, IP marking and IETF ICMP Traceback (ITrace). In this paper [10], we propose an enhancement to the ICMP Traceback approach [11], called ICMP Traceback with Cumulative Path (ITrace-CP). The enhancement consists in encoding the entire attack path information in the ICMP Traceback message. Analytical and simulation studies have been performed to evaluate the performance improvements. We demonstrated that our enhanced solution provides faster construction of the attack graph, with only marginal increase in computation, storage and bandwidth. E. Trace IP Packets by Flexible Deterministic Packet Marking (FDPM) Currently a large number of the notorious Distributed Denial of Service (DDoS) attack incidents make people aware of the importance of the IP traceback technique. IP traceback is the ability to trace the IP packets to their origins. It provides a security system with the capability of identifying the true sources of the attacking IP packets. IP traceback mechanisms have been researched for years, aiming at finding the sources of IP packets quickly and Page 309 precisely. In this paper, an IP traceback scheme, Flexible Deterministic Packet Marking (FDPM) [12], is proposed. It provides more flexible features to trace the IP packets and can obtain better tracing capability over other IP traceback mechanisms, such as link testing, messaging, logging, Probabilistic Packet Marking (PPM) [13] [14], and Deterministic Packet Marking (DPM) [15]. The implementation and evaluation demonstrates that the FDPM needs moderately a small number of packets to complete the traceback process and requires little computation work; therefore this scheme is powerful to trace the IP packets. It can be applied in many security systems, such as DDoS defense systems [4], Intrusion Detection Systems (IDS), forensic systems, and so on.

III. EXISTING SYSTEM Existing IP traceback approaches can be classified into five main categories: packet marking [7] [16], ICMP traceback [11] [10], logging on the router, link testing, overlay, and hybrid tracing. 1) Packet marking methods require routers modify the header of the packet to contain the information of the router and forwarding decision. 2) Different from packet marking methods, ICMP traceback generates addition ICMP messages to a collector or the destination. 3) Attacking path can be reconstructed from log on the router when router makes a record on the packets forwarded. 4) Link testing is an approach which determines the upstream of attacking traffic hop-by-hop while the attack is in progress. 5) Center Track proposes offloading the suspect traffic from edge routers to special tracking routers through a overlay network. V. ADVANTAGES OF PROPOSED SYSTEM 1) This is the first article known which deeply investigates path backscatter messages. These messages are valuable to help understand spoofing activities. Though Moore has exploited backscatter messages, which are generated by the targets of spoofing messages, to study Denial of Services (DoS), path backscatter messages, which are sent by intermediate devices rather than the targets, have not been used in traceback. 2) A practical and effective IP traceback solution based on path backscatter messages, i.e., PIT, is proposed. PIT bypasses the deployment difficulties of existing IP traceback mechanisms and actually is already in force. Though given the limitation that path backscatter messages are not generated with stable possibility, PIT cannot work in all the attacks, but it does work in a number of spoofing activities. At least it may be the most useful traceback mechanism before an AS-level traceback system has been deployed in real. 3) Through applying PIT on the path backscatter dataset, a number of locations of spoofers are captured and presented. Though this is not a complete list, it is the first known list disclosing the locations of spoofers

IV. PROPOSED SYSTEM ARCHITECTURE

The Distributed Denial of Service (DDoS) attacks are launched synchronously from multiple locations and they are extremely harder to detect and stop. Identifying the true origin of the attacker along with the necessary preventive measures helps in blocking further occurrences these types of attacks. The issue of tracing the source of the attack deals with the problem of IP traceback. B. Goals and objectives 1) Designing the IP traceback techniques to disclose the real origin of IP traffic or track the path. 2) A practical and effective IP traceback solution based on path backscatter messages. 3) Passive IP traceback (PIT) that bypasses the deployment difficulties of IP traceback techniques. 4) Packet marking methods to modify the header of the packet to contain the information of the router and forwarding decision. C. Methodologies of Problem Solving And Efficiency Issues: 1) Find the shortest path from source (s) node to destination (d) node. 2) The messassge can be send from r to d through many intermediate nodes i.e. routers (r). 3) There may any spoofer origin available in between the path Assume, that 'sp' is the spoofer node in the network. There are two assumptions for locating such spoofing origin while routing the packets in the network. a) Loop-Free Assumption: This assumption states there is noloop in the paths. This assumption always holds unless misconfiguration or the routing has not converged. b)

Valley-Free Assumption: This assumption states thereshould be no valley in the some node level network paths. Though the increased complexity of node relationship has reduced the universality of this assumption, it is still the most common model of intermediate network level routing. 1) If suppose any intermediate node has being spoofed by spoofer node then the destination node will send the path backscatter message to all intermediate node indicating that spoofing has occurred at somewhere in the network. 2) Then each node in network will send the acknowledgment for that path backscatter message. The node which fails to give back acknowledgment that will be assumed as spoofer node.

V CONCLUSION

In this article we have presented a new technique, backscatter analysis, for estimating denial-of-service attack activity in the Internet. Using this technique, we have observed widespread DoS attacks in the Internet, distributed among many different domains and ISPs. The size and length of the attacks we observe are heavy tailed, with a small number of long attacks constituting a significant fraction of the overall attack volume. Moreover, we see a surprising number of attacks directed at a few foreign countries, at home machines, and towards particular Internet services. We try to dissipate the mist on the the actual locations of spoofers based on investigating the path backscatter messages. In this, we proposed Passive IP Traceback (PIT) which tracks spoofers based on path backscatter messages and public available information. We illustrate causes, collection, and statistical results on path backscatter. We specified how to apply PIT when the topology and routing are both known, or the routing is unknown, or neither of them are known. We presented two effective algorithms to apply PIT in large scale networks and proofed their correctness. We proved that, the effectiveness of PIT based on deduction and simulation. We showed the captured locations of spoofers through applying PIT on the path backscatter dataset.

REFERENCES

[1] C. Labovitz, "Bots, ddos and ground truth," NANOG50, October, vol. 5, 2010.

[2] "The ucsd network telescope."

[3] S. M. Bellovin, "Security problems in the tcp/ip protocol suite," ACM SIGCOMM Computer Communication Review, vol. 19, no. 2, pp. 32–48, 1989.

[4] W. Caelli, S. Raghavan, S. Bhaskar, and J. Georgiades, "Policy and law: denial of service threat," in An Investigation into the Detection and Mitigation of Denial of Service (DoS) Attacks, pp. 41–114, Springer, 2011.

[5] D. Moore, C. Shannon, D. J. Brown, G. M. Voelker, and S. Savage, "Inferring internet denial-of-service activity," ACM Transactions on Computer Systems (TOCS), vol. 24, no. 2, pp. 115–139, 2006.

[6] A. C. Snoeren, C. Partridge, L. A. Sanchez, C. E. Jones, F. Tchakountio, S. T. Kent, and W. T. Strayer, "Hash-based ip traceback," in ACM SIGCOMM Computer Communication Review, vol. 31, pp. 3–14, ACM, 2001.

[7] M. T. Goodrich, "Efficient packet marking for large-scale ip traceback," in Proceedings of the 9th

ACM Conference on Computer and Communications Security, pp. 117–126, ACM, 2002.

[8] S. Savage, D. Wetherall, A. Karlin, and T. Anderson, "Practical network support for ip traceback," in ACM SIGCOMM Computer Communication Review, vol. 30, pp. 295–306, ACM, 2000.

[9] A. Yaar, A. Perrig, and D. Song, "Fit: fast internet traceback," in INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE, vol. 2, pp. 1395– 1406, IEEE, 2005.

[10] H. C. Lee, V. L. Thing, Y. Xu, and M. Ma, "Icmp traceback with cumulative path, an efficient solution for ip traceback," in Information and Communications Security, pp. 124–135, Springer, 2003.

[11] draft-bellovin itrace, "Icmp traceback messages," 2003.

[12] Y. Xiang, W. Zhou, and M. Guo, "Flexible deterministic packet marking: An ip traceback system to find the real source of attacks," Parallel and Distributed Systems, IEEE Transactions on, vol. 20, no. 4, pp. 567–580, 2009. [13] J. Liu, Z.-J. Lee, and Y.-C. Chung, "Dynamic probabilistic packet marking for efficient ip traceback," Computer Networks, vol. 51, no. 3, pp. 866–882, 2007.

[14] M. Adler, "Trade-offs in probabilistic packet marking for ip traceback," Journal of the ACM (JACM), vol. 52, no. 2, pp. 217–244, 2005.

[15] A. Belenky and N. Ansari, "Ip traceback with deterministic packet marking," IEEE communications letters, vol. 7, no. 4, pp. 162–164, 2003.

[16] D. X. Song and A. Perrig, "Advanced and authenticated marking schemes for ip traceback," in INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies.

[17] Proceedings. IEEE, vol. 2, pp. 878–886, IEEE, 2001.

International Journal of Research
Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 14
November 2017

Detection of Cyber bulling Based on the Automatic Code of Marginalized Denoising Improved Semantic

1. KEESARI CHAMANTHI,2. V. RAMESH BABU

1.Pg Scholar, Department Of ECE, Annamacharya Institute Of Technology And Sciences,Piglipur, Batasingaram(V), Hayathnagar(M), Ranga Reddy(D),Hyderabad.

2. Assoc.Professor and Head of the Department, Department Of ECE, Annamacharya Institute Of Technology And Sciences,Piglipur, Batasingaram(V), Hayathnagar(M), Ranga Reddy(D),Hyderabad

ABSTRACT: As a side effect of increasingly popular social media, cyberbullying has emerged as a serious problem afflicting children, adolescents and young adults. Machine learning techniques make automatic detection of bullying messages in social media possible, and this could help to construct a healthy and safe social media environment. In this meaningful research area, one critical issue is robust and discriminative numerical representation learning of text messages. In this paper, we propose a new representation learning method to tackle this problem. Our method named Semantic-Enhanced Marginalized Denoising Auto-Encoder (smSDA) is developed via semantic extension of the popular deep learning model stacked denoising autoencoder. The semantic extension consists ofsemantic dropout noise and sparsity constraints, where the semantic dropout noise is designed based on domain knowledge and the word embedding technique. Our proposed method is able to exploit the hidden feature structure of bullying information and learn a robust and discriminative representation of text. Comprehensive experiments on two public cyberbullying corpora (Twitter and MySpace) are conducted, and the results show that our proposed approaches outperform other baseline text representation learning methods.

INTRODUCTION

SOCIAL Media, as defined in a group of Internet- based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content.'' Via social media, people can enjoy enormous information, convenient communication experience and so on. However, social media may have some side effects such as cyberbullying, which may have negative impacts on the life of people, especially children and teenagers. Cyberbullying can be defined as aggressive, intentional actions performed by an individual or a group of people via digital communication methods such as sending messages and posting comments against a victim. Different from traditional bullying that usually occurs at school during face-to-face communication, cyberbullying on social media can take place anywhere at any time. For bullies, they are free to hurt their peers' feelings because they do not need to face someone and can hide behind the Internet. For victims, they are easily exposed to harassment since all of us, especially youth, are constantly connected to Internet or social media. As reported in cyberbullying victimization rate ranges from 10%to 40%. In the United States, approximately 43% of teenagers were ever bullied on social media . The same as traditional bullying, cyberbullying has negative, insidious and sweeping impacts on children .The outcomes for victims under cyberbullying may even be tragic such as the occurrence of self-injurious behavior or suicides.One way to address the cyberbullying problem is to automatically detect and promptly report bullying messages so that proper measures can be taken to prevent possible tragedies. Previous works on computational studies of bullying have shown that natural language processing and machine learning are powerful tools to study bullying. Cyberbullying detection can be formulated as a supervised learning problem. A classifier is first trained on a cyberbullying corpus labeled by humans, and the learned classifier is then used to recognize a bullying message. Three kinds of information including text, user demography, and social network features are often used in cyberbullying detection. Since the text content is the most reliable, our work here focuses on text-based cyberbullying detection.

In the text-based cyberbullying detection, the first and also critical step is the numerical representation learning for text messages. In fact, representation learning of text is extensively studied in text mining, information retrieval and natural language processing (NLP). Bag-of-words (BoW) model is one commonly used model that each dimension corresponds to a term. Latent Semantic Analysis (LSA) and topic models are another popular text representation models, which are both based on BoW models. By mapping text units into fixed-length vectors, the learned representation can be further processed for numerous language processing tasks. Therefore, the useful representation should discover the meaning behind text units. In cyberbullying detection, the numerical representation for Internet messages should be robust and discriminative. Since messages on social media are often very short and contain a lot of informal language and misspellings, robust representations for these messages are required to reduce their ambiguity. Even worse, the lack of sufficient high-quality training data, i.e., data sparsity make the issue more challenging. Firstly, labeling data is labor intensive and time consuming. Secondly, cyberbullying is hard to describe and judge from a third view due to its intrinsic ambiguities. Thirdly, due to protection of Internet users and privacy issues, only a small portion of messages are left on the Internet, and most bullying posts are deleted. As a result, the trained classifier may not generalize well on testing messages that contain nonactivated but discriminative features. The goal of this present study is to develop methods that can learn robust

and discriminative representations to tackle the above problems in cyberbullying detection. Some approaches have been proposed to tackle these problems by incorporating expert knowledge into feature learning. Yin et.al proposed to combine BoW features, senti- ment features and contextual features to train a support vector machine for online harassment detection [10]. Dinakar et.al utilized label specific features to extend the general features, where the label specific features are learned by Linear Discriminative Analysis [11]. In addition, common sense knowledge was also applied. Nahar et.al presented a weighted TF-IDF scheme via scaling bullying-like features by a factor of two [12]. Besides content-based information, Maral et.al proposed to apply users' information, such as gender and history messages, and context information as extra features . But a major limitation of these approaches is that the learned feature space still relies on the BoW assumption and may not be robust. In addition, the performance of these approaches rely on the quality of hand-crafted features, which require extensive domain knowledge. In this paper, we investigate one deep learning method named stacked denoising autoencoder (SDA) . SDA stacks several denoising autoencoders and concatenates the output of each layer as the learned representation. Each denoising autoencoder in SDA is trained to recover the input data from a corrupted version of it. The input is corrupted by randomly setting some of the input to zero, which is called dropout noise. This denoising process helps the autoencoders to learn robust representation. In addition, each autoencoder layer is intended to learn an increasingly abstract representation of the input . In this paper, we develop a new text representation model based on a variant of SDA:

marginalized stacked denoising autoencoders (mS- DA) which adopts linear instead of nonlinear projection to accelerate training and marginalizes infinite noise distribution in order to learn more robust representations. We utilize semantic information to expand mSDA and develop Semantic-enhanced Marginalized Stacked Denoising Au- toencoders (smSDA). The semantic information consists of bullying words. An automatic extraction of bullying words based on word embeddings is proposed so that the involved human labor can be reduced. During training of smSDA, we attempt to reconstruct bullying features from other normal words by discovering the latent structure, i.e. correlation,

between bullying and normal words. The intuition behind this idea is that some bullying messages do not contain bullying words. The correlation information discovered by smSDA helps to reconstruct bullying features from normal words, and this in turn facilitates detection of bullying messages without containing bullying words. For example, there is a strong correlation between bullying word fuck and normal word off since they often occur together. If bullying

messages do not contain such obvious bullying features, such as fuck is often misspelled as fck, the correlation may help to reconstruct the bullying features from normal ones so that the bullying message can be detected. It should be noted that introducing dropout noise has the effects of enlarging the size of the dataset, including training data size, which helps alleviate the data sparsity problem. In addition, L1 regularization of the projection matrix is added to the objective function of each autoencoder layer in our model to enforce the sparstiy of projection matrix, and this in turn facilitates the discovery of the most relevant terms for reconstructing bullying terms. The main contributions ofour work can be summarized as follows:

* Our proposed Semantic-enhanced Marginalized S- tacked Denoising Autoencoder is able to learn robust features from BoW representation in an efficient and effective way. These robust features are learned by reconstructing original input from corrupted (i.e., missing) ones. The new feature space can improve the performance of cyberbullying detection even with a small labeled training corpus.

* Semantic information is incorporated into the re-construction process via the designing of semantic dropout noises and imposing sparsity constraints on mapping matrix. In our framework, high-quality semantic

information, i.e., bullying words, can be extracted automatically through word embeddings. Finally, these specialized modifications make the new feature space more discriminative and this in turn facilitates bullying detection.

* Comprehensive experiments on real-data sets have verified the performance of our proposed model.

RE LATED WORK

This work aims to learn a robust and discriminative text representation for cyberbullying detection. Text representation and automatic cyberbullying detection are both related to our work. In the following, we briefly review the previous work in these two areas.

TEXT REPRESENTATION LEARNING

In text mining, information retrieval and natural language processing, effective numerical representation of linguistic units is a key issue. The Bag-of-words (BoW) model is the most classical text representation and the cornerstone of some states-of-arts models including Latent Semantic Analysis (LSA) and topic models . BoW model represents a document in a textual corpus using a vector of real numbers indicating the occurrence of words in the document. Although BoW model has proven to be efficient and effective, the representation is often very sparse. To address this problem, LSA applies Singular Value Decomposition (SVD) on the word-document matrix for BoW model to derive a low-rank approximation. Each new feature is a linear combination of all original features to alleviate the sparsity problem. Topic models, including Probabilistic Latent Semantic Analysis and Latent Dirichlet Allocation are also proposed. The basic idea behind topic models is that word choice in a document will be influenced by the topic of the document probabilistically. Topic models try to define the generation process of each word occurred in a document. Similar to the approaches aforementioned, our proposed approach takes the BoW representation as the input. Hoever, our approach has some distinct merits. Firstly, the multilayers and non-linearity of our model can ensure a deep learning architecture for text representation, which has been proven to be effective for learning high-level features. Second, the applied dropout noise can make the learned representation more robust. Third, specific to cyberbullying detection, our method employs the semantic information, including bullying words and sparsity constraint imposed on mapping matrix in each layer and this will in turn produce more discriminative representation.

CYBERBULLYING DETECTION

With the increasing popularity of social media in recent years, cyberbullying has emerged as a serious problem afflicting children and young adults. Previous studies of cyberbullying focused on extensive surveys and its psychological effects on victims, and were mainly conducted by social scientists and psychologists. Although these efforts facilitate our understanding for cyberbullying, the psychological science approach based on personal surveys is very time-consuming and may not be suitable for automatic detection of cyberbullying. Since machine learning is gaining increased popularity in recent years, the computational study of cyberbullying has attracted the interest of researchers. Several research areas including topic detection and affective analysis are closely related to cyberbullying detection. Owing to their efforts, automatic cyberbullying detection is becoming possible. In machine learning-based cyberbullying detection, there are two issues:

1) text representation learning to transform each post/message into a numerical vector and

2) classifier training. Xu et.al presented several off-the-shelf NLP solutions including BoW models, LSA and LDA for representation learning to capture bullying signals in social media.

As an introductory work, they did not develop specialized models for cyberbullying detection. Yin et.al proposed to combine BoW features, sentiment feature and contextual features to train a classifier for detecting possible harassing posts. The introduction of the sentiment and contextual features has been proven to be effective. Dinakar et.al used Linear Discriminative Analysis to learn label specific features and combine them with BoW features to train a classifier [11]. The performance of label-specific features largely depends on the size of training corpus. In addition, they need to construct a bullyspace knowledge base to boost the performance of natural language processing methods. Although the incorporation of knowledge base can achieve a performance improvement, the construction of a complete and general one is labor-consuming. Nahar et.al proposed to scale bullying words by a factor of two in the original BoW features .

The motivation behind this work is quit similar to that of our model to enhance bullying features. However, the scaling operation in is quite arbitrary. Ptaszynski et.al searched sophisticated patterns in a brute-force way. The weights for each extracted pattern need to be calculated based on annotated training corpus, and thus the performance may not be guaranteed if the training corpus has a limited size. Besides content-based information, Maral et.al also employ users' information, such as gender and history messages, and context information as extra features [13],

. Huang et.al also considered social network features to learn the features for cyberbullying detection. The shared deficiency among these for mentioned approaches is constructed text features are still from BoW representation, which has been criticized for its inherent over-sparsity and failure to capture semantic structure . Different from these approaches, our proposed model can learn robust features by reconstructing the original data from corrupted data and introduce semantic corruption noise and sparsity mapping matrix to explore the feature structure which are predictive of the existence of bullying so that the learned representation can be discriminative. Marginalized Denoising Auto-encoder In what follows, we describe our approach. The key idea is to marginalize out the noise of the corrupted inputs in the denoising auto-encoders. We start by describing the conventional denoising auto-encoders and introducing necessary notations. Afterwards, we present the detailed derivations of our approach. Our approach is general and flexible to handle various types of noise and loss functions for denoising. A few concrete examples with popular choices of noise and loss functions are included for illustration. We then analyze the properties of the proposed approach while drawing connections to existing works. 2.1. Denoising Auto-encoder (DAE) The Denoising Auto-Encoder (DAE) is typically implemented as a one-hidden-layer neural network which is trained to reconstruct a data point x ∈ RD from its (partially) corrupted version x˜ (Vincent et al., 2008). The corrupted input x˜ is typically drawn from a conditional distribution p(x˜|x) — common corruption choices are additive Gaussian noise or multiplicative mask-out noise (where values are set to 0 with some probability q and kept unchanged with probability of 1 − q). The corrupted input x˜ is first mapped to a latent representation through the encoder (i.e., the nonlinear transformation between the input layer and the hidden layer). Let z=hθ(x˜) ∈ RDh denote the Dh-dimensional latent representation, collected at the outputs of the hidden layer. The code z is then decoded into the network output y = gθ(z) ∈ RD by the nonlinear mapping from the hidden layer to the output layer. Note that we follow the custom to have both mappings share the same parameter θ. For denoising, we desire y = g ◦ h(x˜) = fθ(x˜) to be as close as possible to the clean data x. To this end, we use a loss function `(x, y) to measure the reconstruction error. Given a dataset D = {x1, · · · , xn}, we optimize the parameter θ by corrupting each xi m-times, yielding x˜ 1 i , . . . , x˜m i , and minimize the averaged reconstruction loss 1 n Xn i=1 1 m Xm j=1 ` xi , fθ(x˜ j i)) . (1) Typical choices for the loss ` are the squared loss for realvalued inputs, or the cross-entropy loss for binary inputs

CONCLUSION

This paper addresses the text-based cyberbullying detection problem, where robust and discriminative representations of messages are critical for an effective detection system. By designing semantic dropout noise and enforcing sparsity, we have developed semantic-enhanced marginalized denoising autoencoder as a specialized representation learning model for cyberbullying detection. In addition, word embeddings have been used to automatically expand and refine bullying word lists that is initialized by domain knowledge. The performance of our approaches has been experimentally verified through two cyberbullying corpora from social medias: Twitter and MySpace. As a next step we are planning to further improve the robustness of the Term Reconstruction on Twitter datasets. Each Row Shows Specific Bullying Word, along with Top-4 Reconstructed Words (ranked with their frequency values from top to bottom) via mSDA (left column) and smSDA (right column).

REFERENCES

[1] A. M. Kaplan and M. Haenlein, “Users of the world, unite! The challenges and opportunities of social media,” Business horizons, vol. 53, no. 1, pp. 59–68, 2010.

[2] R. M. Kowalski, G. W. Giumetti, A. N. Schroeder, and M. R. Lattanner, “Bullying in the digital age: A critical review and meta- analysis of cyberbullying research among youth.” 2014.

[3] M. Ybarra, “Trends in technology-based sexual and non-sexual aggression over time and linkages to nontechnology aggression,” National Summit on Interpersonal Violence

and Abuse Across the Lifespan: Forging a Shared Agenda, 2010.

[4] B. K. Biggs, J. M. Nelson, and M. L. Sampilo, "Peer relations in the anxiety–depression link: Test of a mediation model," Anxiety, Stress, & Coping, vol. 23, no. 4, pp. 431–447, 2010.

[5] S. R. Jimerson, S. M. Swearer, and D. L. Espelage, Handbook of bullying in schools: An international perspective. Routledge/Taylor & Francis Group, 2010.

[6] G. Gini and T. Pozzoli, "Association between bullying and psychosomatic problems: A meta-analysis," Pediatrics, vol. 123, no. 3, pp. 1059–1065, 2009.

[7] A. Kontostathis, L. Edwards, and A. Leatherman, "Text mining and cybercrime," Text Mining: Applications and Theory. John Wiley & Sons, Ltd, Chichester, UK, 2010.

[8] J.-M. Xu, K.-S. Jun, X. Zhu, and A. Bellmore, "Learning from bullying traces in social media," in Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies. Association for Computational Linguistics, 2012, pp. 656–666.

[9] Q. Huang, V. K. Singh, and P. K. Atrey, "Cyber bullying detection using social and textual analysis," in Proceedings of the 3rd Inter- national Workshop on Socially-Aware Multimedia. ACM, 2014, pp. 3–6.

[10] D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, and L. Edwards, "Detection of harassment on web 2.0," Proceedings of the Content Analysis in the WEB, vol. 2, pp. 1–7, 2009.

[11] K. Dinakar, R. Reichart, and H. Lieberman, "Modeling the detec- tion of textual cyberbullying." in The Social Mobile Web, 2011.

[12] V. Nahar, X. Li, and C. Pang, "An effective approach for cy- berbullying detection," Communications in Information Science and Management Engineering, 2012.

[13] M. Dadvar, F. de Jong, R. Ordelman, and R. Trieschnigg, "Im- proved cyberbullying detection using gender information," in Proceedings of the 12th - Dutch-Belgian Information Retrieval Workshop (DIR2012). Ghent, Belgium: ACM, 2012.

[14] M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, "Im- proving cyberbullying detection with user context," in Advances in Information Retrieval. Springer, 2013, pp. 693–696.

[15] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful

representations in a deep network with a local denoising criterion," The Journal of Machine Learning Research, vol. 11, pp. 3371–3408, 2010.

[16] P. Baldi, "Autoencoders, unsupervised learning, and deep archi- tectures," Unsupervised and Transfer Learning Challenges in Machine Learning, Volume 7, p. 43, 2012.

[17] M. Chen, Z. Xu, K. Weinberger, and F. Sha, "Marginalized de- noising autoencoders for domain adaptation," arXiv preprint arX- iv:1206.4683, 2012.

[18] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," Discourse processes, vol. 25, no. 2-3, pp. 259–284, 199

International Journal of Research
Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017

Exploiting Online Social Behaviors for Compromised Account Detection

Addada Sai Akshitha Devi & Ramesh Varugu

[1],Pg Scholar , [2]Asst. Professor , Department Of Cse, Annamacharya Institute Of Technology And Sciences Piglipur, Batasingaram, Hayat Nagar, Ranga Reddy District, Hyderabad, Telangana 501512

Abstract—*A social behavioral profile precisely mirrors a client's OSN action designs. Individuals get to OSNs utilizing both customary desktop PCs and new developing cell phones. With more than one billion clients around the world, OSNs are another scene of development with many testing research issues. In this paper, we contemplate the social practices of OSN clients, i.e., their utilization of OSN administrations, and the use of which in recognizing the traded off records. We catch client conduct with the accompanying measurements: client availability, client movement and client reactions.we approve and portray the client social action on OSN.The consider depends on nitty gritty clickstream information ,the clickstream information uncovers key highlights of the interpersonal organization workloads, for example, how every now and again individuals associate with informal communities and for to what extent, and in addition the sorts and arrangements of exercises that clients lead on these locales. we focus on the qualities of social practices we survey malevolent practices of OSN clients and demonstrate the social behavioral profiles can precisely separate individual OSN clients and recognize traded off records.*

I Introduction

Online informal communities (OSNs) have moved toward becoming to a great degree popular.Social media have pulled in front of email as the most mainstream online action.

More than 66% of the worldwide online populace visit and take an interest in interpersonal organizations and sites. Indeed, long range informal communication and blogging represent about 10% ever spent on the Internet. These insights propose that OSNs have turned into a basic piece of the worldwide online experience. OSN client conduct covers different social exercises that clients can do on the web, for example, companionship creation, content distributing, profile perusing, informing, and remarking. Quite, these exercises can be true blue or vindictive. Seeing how clients act when they interface with these locales is critical for various reasons in light of the fact that nowadays traded off records are focused on or we can state favored by spammers. The noxious one breaks the put stock seeing someone between the honest to goodness account proprietors and their companions, and productively circulate spam advertisements, phishing joins, or malware. Presently a day's hacking somebody's online person to person communication profiling traits and after that utilization of the same for any obscene exercises is been a genuine risk. The record of famous people or political pioneers is for the most part trap for this sort of framework. Numerous frameworks are been proposed to distinguish this sort of profiling assault yet the majority of them are transfer on watched realities about the bookkeeping which by and large sets aside longer opportunity to recognize the assault. So to reduce this season of discovery for beginning times of the bargained assaults framework ought to have equipped for recognition of concealed states. This thought in the long run expands the early identification which can maintain a strategic distance from genuine dangers.

II Related Work

[1] Proposed System distinguishes Towards Detecting Compromised Accounts on Social Networks. Proposed strategies aid to distinguish and avoid three true assaults against well known organizations and news offices Attacker who knows about COMPA can keep account from location. Mechanized creeping backing off such information gathering attempts. COMPA can be effortlessly stretched out with extra and more Comprehensive similitude measures. Future extension: Other comparability measures reconciliation is extent of work. [2] This article presents investigative work on how clients' movement on Face book identifies with their identity, as measured by the standard Five Factor Model. Results indicate noteworthy connections between identity characteristics and different highlights of Face

book profiles. We at that point demonstrate how multivariate relapse permits forecast of the identity attributes of an individual client given their Facebook profile. Impediments: Data utilized may experience the ill effects of a self-choice inclination. clients could control the data put away with respect to their profile, so we just had information for clients who let us get to this data. Future degree: Online publicizing and recommender frameworks [3] Website entryway concentrates on long range interpersonal communication accounts that have been hacked and appraises that 50000 watchword and utilize free been stolen. Impediments: Privacy and security on informal community is in question. Future degree: Better protection and security in interpersonal organizations. [4] Present new kind investigation of client workloads in online interpersonal organizations. Our examination depends on itemized click stream information, gathered over a 12-day time span, compressing HTTP sessions of 37,024 clients who got to four famous interpersonal organizations: Orkut, MySpace, Hi5, and LinkedIn. Investigation exhibits energy of utilizing click stream information in distinguishing designs in interpersonal organization workloads and social communications. Perusing, which can't be gathered from creeping openly accessible information, represents 92% of client exercises. Constraint: Huge information is been crept and henceforth require better calculations for improved handling. Future Scope: Future examination work is discovering Impact of companions on conduct of client of informal communities. Inspired by understanding substance dispersion designs over numerous OSNs.interpersonal organization workload generator and Markov models.

III System Overview

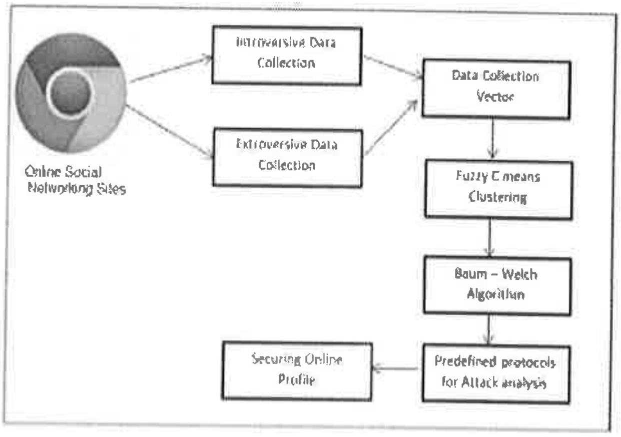

Online Social Networking Sites
Introversive Data Collection
Extroversive Data Collection
Data Collection Vector
Fuzzy C means Clustering
Baum – Welch Algorithm
Predefined protocols for Attack analysis
Securing Online Profile

Figure 1. System overview

Social Behavior Features: We categorize user social behaviors on an OSN into two classes, extroversive behaviors and introversive behaviors. Extroversive behaviors, such as uploading photos and sending messages, result in visible imprints to one or more peer users; introversive behaviors, such as browsing other users' profiles and searching in message inbox, however, do not produce observable effects to other users. While most previous research only focus on the extroversive behaviors, such as public posting [8], we study both classes of behaviors for a more complete understanding and characterization of user social behaviors.

A. Extroversive Behavior Features: Extroversive Behaviors directly reflect how a user interacts with its friends online, and thus they are important for characterizing a user's social behaviors.

B. Introversive Behavior Features: Although invisible to peer users, introversive behaviors make up the majority of a user's OSN activity; as studied in previous work [6], [15] the dominant (i.e., over 90%) user behavior on an OSN is browsing. Through introversive activities users gather and consume social information, which helps them to form ideas and opinions, and eventually, establish social connections and initiate future social communications. Hence, introversive behavior patterns make up an essential part of a user's online social behavioral characteristics. We propose the following four features to portray a user's introversive behavior. 1.Fuzzy

C- Means Clustering : Here in this step all the data collected in above two steps are been formatted and collected in a list. And then this list is been subjected to labeling of the entities for numerical conversions and then based on this data is been converted into clusters of the desires facts by using Fuzzy C means process. Here all the data that is been collected for the calming of insurance is clustered logically using c means clustering with the following technique. This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one. After each iteration membership and cluster centers are updated according to the formula. 2.Baum_

Welch Methodology : The Baum Welch algorithm is used to extract the hidden states from the k known parameters like introversive and extroversive entities and Baum Welch algorithm is mentioned below. Baum- Welch Algorithm

Input : Data Set D, Observed States Os = { Os1 , Os2,Os3}

Step 0: Start

Step 1: Identify the Observed state Attribute Osi

Step 2: FOR i=0 to size of D

Step 3: Identify Attribute Osi and put in separateList OSL

Step 4: END FOR

Step 5: Transaction count Tc=0

Step 6: FOR i=0 to size of OSL

Step 7: identify a and ß

Step 8: Compute using Equation1

Step 9:IF belongs to Os

Step 10: THEN add Hs (Hidden State) to list

Step 11: END FOR

Step 12: Stop

IV Conclusion

In this paper, we propose to fabricate a social conduct profile for individual OSN clients to describe their behavioral examples. Our approach considers both extroversive and introversive practices. In this paper, we propose to assemble a social conduct profile for individual OSN clients to describe their behavioral examples. Our approach considers both extroversive and introversive behaviors.Proper distinguishing proof of shrouded conditions of assault.

References

[1]M. Egele, G. Stringhini, C. Kruegel, and G. Vigna, "COMPA: Detecting compromised accounts on social networks"in Proc. Symp. Netw.Distrib. Syst. Secur. (NDSS), San Diego, CA, USA, 2013.

[2]F. Schneider, A. Feldmann, B. Krishnamurthy, and W.Willinger, "Understanding online social network usage from a network perspective,"Proc. 9th ACM SIGCOMM Conf. Internet Meas. Conf. (IMC), Chicago,IL, USA, 2009, pp. 35–48.

[3]G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks,"in Proc. 26th Annu. Comput.Secur. Appl. Conf. (ACSAC),Austin, TX, USA, 2010, pp. 1–9.

[4]FabrícioBenevenutoy Tiago RodriguesyMeeyoungChalVirgílio Almeida, "Characterizing User Behavior in Online Social Networks" in Proc. 9th ACM SIGCOMM Conf. Internet Meas. Conf. (IMC), Chicago, IL, USA, 2009, pp. 49–62

. [5] Tiago Rodrigues Meeyoung Cha Virgílio Almeida "Characterizing User Behaviorin Online Social Networks" in proc13

[6] Y. Xieet al., "Innocent by association: Early recognition of legitimate users," in Proc. ACM Conf. Comput. Commun.Secur.(CCS), Raleigh, NC, USA, 2012, pp. 353–364

[7] H. Xiong, P. Malhotra, D. Stefan, C. Wu, and D. Yao, "User-assisted host-based detection of outbound malware traffic," in Proc. 11th Int.Conf. Inf. Commun. Secur. (ICICS), Beijing, China, 2009, pp. 293–307.

[8] C. Yang, R. Harkreader, J. Zhang, S. Shin, and G. Gu, "Analyzing spammers' social networks for fun and profit: A case study of cyber criminal ecosystem on Twitter," in Proc. 21st Int. Conf. World WideWeb (WWW), Lyon, France, 2012, pp. 71–80.

[9] C. Yang, R. C. Harkreader, and G. Gu, "Die free or live hard? Empirical evaluation and new design for fighting evolving Twitter spammers," in Proc. 14th Int. Conf. Recent Adv. Intrusion Detection (RAID), Menlo Park, CA, 2011, pp. 318–337.

EXPLORE THE MALICIOUS FACEBOOK APPLICATIONS

S.Komali

Assistant Professor, Department of CSE, Dhruva Institute of Technology & Sciences (India)

ABSTRACT

Applications that present suitable means for hackers to spread malicious content on Face book on the other hand, little is understood concerning features of malicious applications and how they function. In the recent times, hackers have considered popularity of the third-party application platform as well as deployment of malicious applications. Our aim is to build up a rigorous application evaluator of face book which is the first tool that is focused on detection of malicious applications on Face book. There are lots of malicious applications spreading on Face book each day. This is possibly initial comprehensive study that has focused on malicious Face book applications that focus on quantifying as well as understanding of malicious applications and make this information into an effectual detection method. For structuring of rigorous application evaluator of face book, we make use of data from a security application within Face book that examines profiles of Face book users. To build up rigorous application evaluator of face book we make use of information which is gathered by means of observation of posting behaviour of Face book apps which are seen across millions of face book users.

Keywords: Malicious applications, Face book, Third-party application, Security.

I. INTRODUCTION

The research community has paid less consideration towards social networking applications up to now. Most of the research which is associated to spam and malware on Face book has spotlighted on detection of malicious posts as well as social spam operations. Simultaneously, in apparently backwards move, Face book has dismantled its application rating in recent times [1]. There are several means that hackers can advantage from malicious app such as: the application reaching huge numbers of users as well as their friends to extend spam; the application obtains user personal data; application reproduces by making other applications acceptable means. To make matter severe, usage of malicious applications is cut down by ready-to-use toolkits. Applications of third-party are the most important reason for popularity as well as addictiveness of Face book. Unfortunately, hackers have understood potential of usage of applications for spreading of malware as well as spam. Usage of huge corpus of malicious face book applications show that malicious applications change from benign applications regarding numerous features. In the recent times, a user has extremely restricted information during the time of installing an application on Face book. When provided an application identity number, we can detect when an application is malicious or not. In the recent times, there is no commercial service, openly available information to give advice a user regarding the risks of an application [2]. For structuring of rigorous application evaluator of face book, we make use of data from a security application within Face book that examines profiles of Face book users. The proposed system identifies malicious applications by means of using only features that are obtained on-demand or usage of on-demand as well as aggregation-based application data. Our aim is to develop a rigorous application evaluator of face book which is the first tool that is focused on detection of malicious applications on Face book. To develop rigorous application evaluator of face book we make use of information which is gathered by means of observation of posting behaviour of Face book apps which are seen across millions of face book users.

II. METHODOLOGY

For building of rigorous application evaluator of face book, we make use of data from MyPage-Keeper which is a security application within Face book that examines profiles of Face book users. This is perhaps the initial comprehensive study that has focused on malicious Face book applications that focus on quantifying as well as understanding of malicious applications and make this information into an effectual detection method. Online social networks will permit applications of third-party to enhance user experience above these platforms. To develop rigorous application evaluator of face book we make use of information which is gathered by means of observation of posting behaviour of Face book apps which are seen across millions of face book users. Driving motivation for detection of malicious applications will develop from suspicion that important fraction of malicious posts on Face book are posted by means of applications. We develop a rigorous application evaluator of face book which is the first tool that is focused on detection of malicious applications on Face book [3]. In our work usage of huge corpus of malicious face book applications which are observed show that malicious applications change from benign applications regarding numerous features. Long term, we observe rigorous application evaluator of face book as a move towards creation of independent watchdog for assessment as well as ranking of applications, in order to advise Face book users earlier than installing of applications [4]. These improvements include interesting means of communicating between online friends as well as various activities. Initially we distinguish several features that assist us in differentiation of malicious applications from the benign ones. Secondly, leveraging these distinctive features, the proposed rigorous application evaluator of face book will identify malicious applications with more accuracy, with no false positives. There is a number of community based feedback motivated efforts to grade applications although these might be extremely powerful in future, up to now they have received little acceptance.

III. AN OVERVIEW OF PROPOSED SYSTEM

Unlike distinctive desktop as well as smart phone applications, installation of application by user does not include user downloading and execution of application binary. Whenever a user adds Face book application to their profile, user provides application server permission towards subset of data which is listed on user Face book profile and permission to carry out assured actions in aid of user. After that, application can have access to data and carry out legalized actions in support of user. So far, research studies got focused on detection of malicious posts as well as campaigns. We develop a rigorous application evaluator of face book which is the first tool that is focused on detection of malicious applications on Face book. In the third step, application afterwards access personal data from user profile, which hackers potentially make use of to profit. In the step four, application makes malicious posts in support of user to lure user friends to set up the similar application and by this means the cycle will continues with application or else colluding applications reaching more users. To develop rigorous application evaluator of face book we make use of information which is gathered by means of observation of posting behaviour of Face book applications which are seen across millions of face book users. On the other hand, even the early work leads to recommendations in support of Face book that might be useful for other social platforms. Face book permits third-party developers to present services towards its users by Face book applications [5]. It acts as a move towards creation of independent watchdog for assessment as well as ranking of applications, in order to advise Face book users earlier than installing of applications. In the fig1 showing operations of Facebook application, includes several steps. This is possibly the first comprehensive work that has focused on malicious Face book applications that focus on quantifying as well as understanding of malicious applications and make this information into an effectual detection method. Proposed evaluator of face book will identify malicious applications with more accuracy, with no false positives. In the initial step, hackers convince users to set up the application, typically with some false promise. In the other step, when a user set up the application, it redirects user towards web page in which user is appealed to carry out tasks [6]. The proposed rigorous application evaluator of face book identifies malicious applications by means of using only features that are obtained on-demand or usage of on-demand as well as aggregation-based application data. Important message of our work is that there looks to be parasitic eco-system of malicious applications in Face book that requires be stopping.

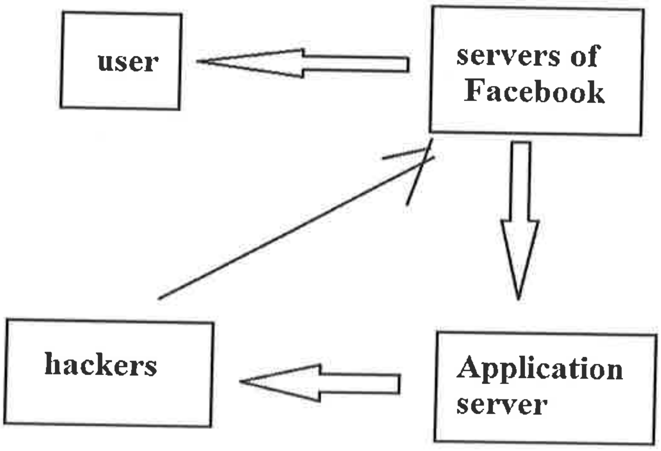


Fig1: Operation process of a Face book application

IV. CONCLUSION

This is possibly initial comprehensive study that has focused on malicious Face book applications that focus on quantifying as well as understanding of malicious applications and make this information into an effectual detection method. To build up thorough application evaluator of face book we make use of information which is gathered by means of observation of posting behaviour of Face book apps which are seen across millions of face book users. The recent works studies regarding application permissions and how community ratings associate to privacy threats of Face book applications. We build up a rigorous application evaluator of face book which is the first tool that is focused on detection of malicious applications on Face book. The projected rigorous application evaluator of face book identifies malicious applications by means of using only features that are obtained on-demand or usage of on-demand as well as aggregation-based application data. We study proposed rigorous application evaluator of face book as a move towards creation of independent watchdog for assessment as well as ranking of applications, in order to advise Face book users earlier than installing of applications.

REFERENCES

[1] Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Analyzing facebook privacy settings: user expectations vs. reality. In IMC, 2011.

[2] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker. Beyond blacklists: learning to detect malicious web sites from suspicious urls. In KDD, 2009.

[3] A. Makridakis, E. Athanasopoulos, S. Antonatos, D. Antoniades, S. Ioannidis, and E. P. Markatos. Understanding the behavior of malicious applications in social networks. Netwrk. Mag. of Global Internetwkg., 2010.

[4] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song. Design and Evaluation of a Real-Time URL Spam Filtering S ervice. In Proceedings o the IEEE Symposium on Security and Privacy, 2011.

[5] N. Wang, H. Xu, and J. Grossklags. Third-party apps on facebook: privacy and the illusion of control. In CHIMIT, 2011.

[6] C. Yang, R. Harkreader, and G. Gu. Die free or live hard? empirical evaluation and new design for fighting evolving twitter spammers. In RAID, 2011.

IJITECH International Journal of Innovative Technologies
WWW.IJITECH.ORG

ISSN 2321-8665
Vol.04,Issue.13,
September-2016,
Pages:2323-2329

Enabling Fine-Grained Multi-Keyword Search Supporting Classified Sub-Dictionaries over Encrypted Cloud Data

G.SANDHYA RANI[1], RAMESH BABU VARUGU[2], V.VEDASAHITHI[3]

[1]PG Scholar, Dept of CSE, Annamacharya Institute of Technology and Science, Hyderabad, TS, India,
E-mail: rockzsandhya@gmail.com.

[2]Associate Professor& HOD, Dept of CSE, Annamacharya Institute of Technology and Science, Hyderabad, TS, India,
E-mail: ramesh.vnl@gmail.com.

[3]Assistant Professor, Dept of CSE, Annamacharya Institute of Technology and Science, Hyderabad, TS, India,
E-mail: sahithivellanki@gmail.com.

Abstract: Using cloud computing, individuals can store their data on remote servers and allow data access to public users through the cloud servers. As the outsourced data are likely to contain sensitive privacy information, they are typically encrypted before uploaded to the cloud. This, however, significantly limits the usability of outsourced data due to the difficulty of searching over the encrypted data. In this paper, we address this issue by developing the fine-grained multi-keyword search schemes over encrypted cloud data. Our original contributions are three-fold. First, we introduce the relevance scores and preference factors upon keywords which enable the precise keyword search and personalized user experience. Second, we develop a practical and very efficient multi-keyword search scheme. The proposed scheme can support complicated logic search the mixed "AND", "OR" and "NO" operations of keywords. Third, we further employ the classified sub-dictionaries technique to achieve better efficiency on index building, trapdoor generating and query. Lastly, we analyze the security of the proposed schemes in terms of confidentiality of documents, privacy protection of index and trapdoor, and unlinkability of trapdoor. Through extensive experiments using the real-world dataset, we validate the performance of the proposed schemes. Both the security analysis and experimental results demonstrate that the proposed schemes can achieve the same security level comparing to the existing ones and better performance in terms of functionality, query complexity and efficiency.

Keywords: Searchable Encryption, Multi-Keyword, Fine-Grained, Cloud Computing.

I. INTRODUCTION

Transmitting the information to the cloud servers the data encryption, though, would considerably lower the usability of data outstanding to the complexity of penetrating over the encrypted data purely encrypting the statistics may still basis other sanctuary concerns. For example, Google Search uses SSL (Secure Sockets Layer) to encrypt the association among search user and Google server when confidential data, such as credentials and emails, appear in the search results. Nevertheless, if the explore user clicks into a different website as of the search consequences page, that website may be talented to categorize the explore terms that the user has worn. Firstly, the statistics owner needs to produce numerous keywords according to the outsourced data. These keywords are then encrypted and stored at the cloud server. When a explore user requirements to admission the outsourced data, it can select some appropriate keywords and send the nothing text of the preferred keywords to the cloud server. The cloud server then uses the cipher text to match the outsourced encrypted keywords, and lastly returns the matching results to the search user. To achieve the similar search efficiency and precision over encrypted data as that of plaintext keyword search, an extensive body of research has been developed in literature. Propose a multi-keyword text search scheme which considers the relevance scores of keywords and utilizes a multidimensional tree technique to achieve efficient search query.

Yu et al. propose a multi-keyword top-k retrieval scheme which uses fully homomorphism encryption to encrypt the index/trapdoor and guarantees high security. Cao et al. propose a multi-keyword ranked search (MRSE), which applies coordinate machine as the keyword matching rule, i.e., return data with the most matching keywords. Although many search functionalities have been developed in previous literature towards precise and efficient searchable encryption, it is still difficult for searchable encryption to achieve the same user experience as that of the plaintext search, like Google search. The relevance scores of keywords can enable more precise returned results, and the preference factors of keywords represent the importance of keywords in the search keyword set specified by search users and correspondingly enables personalized search to cater to specific user preferences. It thus further improves the search functionalities and user experience.

II. EXISTING AND PROPOSED SYSTEMS

A. Existing System

The searchable encryption has been recently developed as a fundamental approach to enable searching over encrypted cloud data, which precedes the following operations. Wang et al. propose a ranked keyword search scheme which considers the relevance scores of keywords. Sun et al. propose a multi-keyword text search scheme which considers the relevance scores of keywords and utilizes a multidimensional tree technique to achieve efficient search query. Yu et al. propose a multi-keyword top-k retrieval scheme which uses fully homomorphic encryption to encrypt the index/trapdoor and guarantees high security. Cao et al. propose a multi-keyword ranked search (MRSE), which applies coordinate machine as the keyword matching rule, i.e., return data with the most matching keywords.

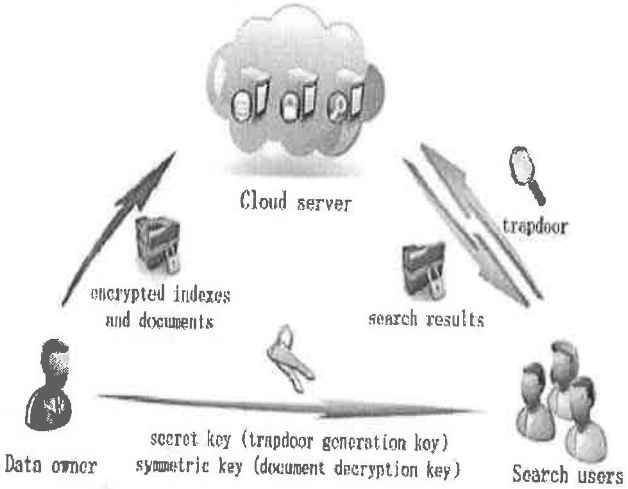

Cloud server
trapdoor
encrypted indexes and documents
search results
secret key (trapdoor generation key)
symmetric key (document decryption key)
Data owner
Search users

Fig.1. System Architecture

B. Proposed System

In this work, we address by developing two Fine-grained Multi-keyword Search (FMS) schemes over encrypted cloud data. In this system, we introduce the relevance scores and the preference factors of keywords for searchable encryption. The relevance scores of keywords can enable more precise returned results, and the preference factors of keywords represent the importance of keywords in the search keyword set specified by search users and correspondingly enables personalized search to cater to specific user preferences. It thus further improves the search functionalities and user experience. In this system, we realize the "AND", "OR" and "NO" operations in the multi-keyword search for searchable encryption. Compared with schemes, the proposed scheme can achieve more comprehensive functionality and lower query complexity. In this system, we employ the classified sub-dictionaries technique to enhance the efficiency of the above two schemes. Extensive experiments demonstrate that the enhanced schemes can achieve better efficiency in terms of index building, trapdoor generating and query in the comparison with schemes

1. Advantages of Proposed System

- Better search results with multi-keyword query by the cloud server according to some ranking criteria.
- To reduce the communication cost.
- Achieves lower query complexity.
- Achieves better efficiency in index building scheme of our proposed model.

III. MODULE DESCRIPTION

A. Searchable Encryption

This module used on search the key word for encrypted text. This Module Used On Secure Purpose. More Secure for Encrypted Encryption to achieve the same user experience as that of the plaintext search, like Google search. This mainly attributes to following two issues. Firstly, query with user preferences is very popular in the Chipper text search

B. Multi-Keyword

Text search scheme which considers the relevance scores of keywords and utilizes a multidimensional tree technique to achieve efficient search query Multi keyword top-k retrieval scheme which uses fully homomorphic encryption to encrypt the index/trapdoor and guarantees high security Cao et al, propose a multi-keyword ranked search (MRSE), which applies coordinate machine as the keyword matching rule, i.e., return data with the most matching keywords Fine-grained Multi-keyword Search (FMS) schemes over encrypted cloud data. Multi-keyword search and coordinate matching using secure kNN computation scheme multi-keyword top-k retrieval scheme with fully homomorphic encryption, which can return ranked results and achieve high security.

C. Fine-Grained

We propose FMS (CS) schemes which not only support multi-keyword search over encrypted data, but also achieve the fine-grained keyword search with the function to investigate the relevance scores and the preference factors of keywords and, more importantly the logical rule of keywords. In addition, with the classified sub-dictionaries, our proposal is efficient in terms of index building, trapdoor generating and query fine-grained operations of keyword search, i.e., "AND", "OR" and "NO" operations in Google Search, which are definitely practical and significantly enhance the functionalities of encrypted keyword search.

D. Cloud Computing

Cloud computing is a computing term or metaphor that evolved in the late 1900s, based on utility and consumption of computer resources. Cloud computing involves deploying groups of remote servers and software networks that allow different kinds of data sources be uploaded for real time processing to generate computing results without the need to store processed data on the cloud. Clouds can be classified as public, private or hybrid. Synonym expansions are words with the same or similar meanings. In order to improve the accuracy of search results, the A Secure and Dynamic Multi-keyword Ranked extracted from out sourced text documents need to be extended by common synonyms, as cloud customers' searching input might be the synonyms of the predefined A Secure and Dynamic Multi-keyword Ranked, not the exact or fuzzy matching A Secure and Dynamic Multi-keyword Ranked due to the possible synonym substitution and/or her lack of exact knowledge about the data. A common synonym thesaurus is built on the foundation of the New American Roget's College Thesaurus (NARCT). Then the keyword set is extended by using the constructed synonym thesaurus. Cryptography The art of protecting information by transforming it (encrypting it) into an unreadable format, called cipher text only those who possess a secret key can decipher (or decrypt) the message into plain text. Encrypted messages can sometimes be broken by cryptanalysis, also called code breaking, although modern cryptography techniques are virtually unbreakable.

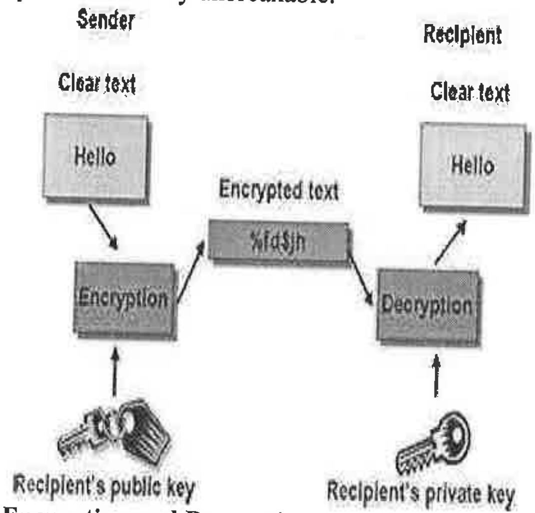


Fig.2. Encryption and Decryption

1. Encryption: In an encryption scheme, the message or information (referred to as *plaintext*) is encrypted using an encryption algorithm, turning it into an unreadable *cipher text* (ibid.). This is usually done with the use of an *encryption key,* which specifies how the message is to be encoded. Any adversary that can see the cipher text, should not be able to determine anything about the original message.

2. Decryption*:* An authorized party, however, is able to decode the cipher text using a decryption algorithm that usually requires a secret decryption key that adversaries do not have access to. For technical reasons, an encryption scheme usually needs a key-generation algorithm, to randomly produce keys. Hence, it is an especially important thing to explore an effective multi-keyword ranked searching service over encrypted outsourced data.

IV. PERFORMANCE EVALUATIONS

In this section, we evaluate the performance of the proposed schemes using simulations, and compare the performance with that of existing proposals in [7]. We apply a real world dataset from the National Science Foundation Research Awards Abstracts 1990-2003, in which we random select multiple documents and conduct real-world experiments on an Intel Core i5 3.2 GHz system.

A. Functionality

We compare functionalities between [7] and our schemes in Table 1, where I and II represent FMS (CS) I and FMS (CS) II, respectively. MRSE [7] can achieve multi-keyword search and coordinate matching using secure kNN computation scheme. And considers the relevance scores of keywords. Compared with the other schemes, our FMS (CS) I consider both the relevance scores and the preference factors of keywords. Note that if the search user sets all relevance scores and preference factors of keywords as the same, the FMS (CS) I degrade to MRSE and the coordinate matching can be achieved. And in the FMS (CS) II, if the search user sets all preference factors of "OR" operation keywords as the same, the FMS (CS) II can also achieve the coordinate matching of "OR" operation keywords particularly, the FMS (CS) II achieves some fine-grained operations of keyword search, i.e., "AND", "OR" and "NO" operations in Google Search, which are definitely practical and significantly enhance the functionalities of encrypted keyword search.

TABLE I. Comparison of Functionalities

	[6]	[13]	[14]	I	II
Multi-keyword search	√	√	√	√	√
Coordinate matching	√	√	√	√	√
Relevance score		√	√	√	
Preference factor				√	√
AND OR NO operations					√

B. Query Complexity

In the FMS (CS) II, we can implement "OR", "AND" and "NO" operations by defining appropriate weights of keywords, this scheme provides a more fine-grained search than [7]. If the keywords to perform "OR", "AND" and "NO" operations are $(w'_1, w'_{2,} \cdots ;w'_{l1})$, $(w''_1 ;w''_{2,} \cdots ;w''_{l2})$and $(w'''_1, w'''_2, \cdots ,w''_{l3})$, respectively. Our FMS (CS) II can complete the search with only one query, however, in [7]; they would complete the search through the following steps:

- For the "OR" operation of l_1 keywords, they need only one query *Query* $(w'_1, w'_2, \cdots ;w'_{l1})$ to return a collection of documents with the most matching keywords (i.e., coordinate matching), which can be denoted as $X = Query(w'_1, w'_2, \cdots ,w'_{l1})$.
- For the "AND" operation of l_2 keywords, [7] cannot generate a query for multiple keywords to achieve the "AND" operation. Therefore, after costing l_2 queries *Query* $(w''_i)(i = 1, 2, \cdots, l_2)$, they can do the "AND" operation, and the corresponding document set can be denotedas $Y=Query(w''_1) \cap Query(w''_2) \cap \cdots \cap Query(w''_{l2})$.
- For the "NO" operation of l_3 keywords, they need l_3 queries *Query* (w'''_i) $(i = 1, 2, \cdots ,l_3)$, firstly. Then, the document set of the "NO" operation can be denoted as $Z = Query(w'''_1) \cap Query(w'''_2) \cap \cdots \cap Query(w'''_{l3})$.
- Finally, the document collection achieved "OR", "AND" and "NO" operations can be represented as $X \cap Y \cap Z$.

As shown in Fig. 3a, 3b and 3c, to achieve these operations, the FMS (CS) II can outperform the existing proposals with less queries generated.

C. Efficiency

1. Computation Overhead

In order to easily demonstrate our scheme computation overhead, we analysis our scheme from each phase index building note that the *Index building* phase of [7] is the same as our FMS II scheme, without calculating the relevance score. And the *Index building* phase of the FMS I is the same as, containing the relevance score computing. Compared with the FMS I, the FMS II do not need to calculate the relevance score. And compared with the computation cost of building index, the cost of calculating the relevance score is negligible, we do not distinguish them. Moreover, in our enhanced schemes (FMSCS), we divide the total dictionary into 1 common sub-dictionary and 20 professional sub-dictionaries (assume each data owner averagely chooses 1 common sub-dictionary and 3 professional sub-dictionaries to generate the index). As shown in Fig. 4, we can see the time for building index is dominated by both the size of dictionary and the number of documents. And compared with [7], and our FMS schemes, the FMSCS schemes largely reduce the computation overhead. Trapdoor generating: In Trapdoor generating phase, [7] firstly creates a vector according to the search keyword set$\widetilde{W}$, then encrypts the vector by the secure kNN computation scheme. And also generates a vector and uses homomorphic encryption to encrypt each dimension.

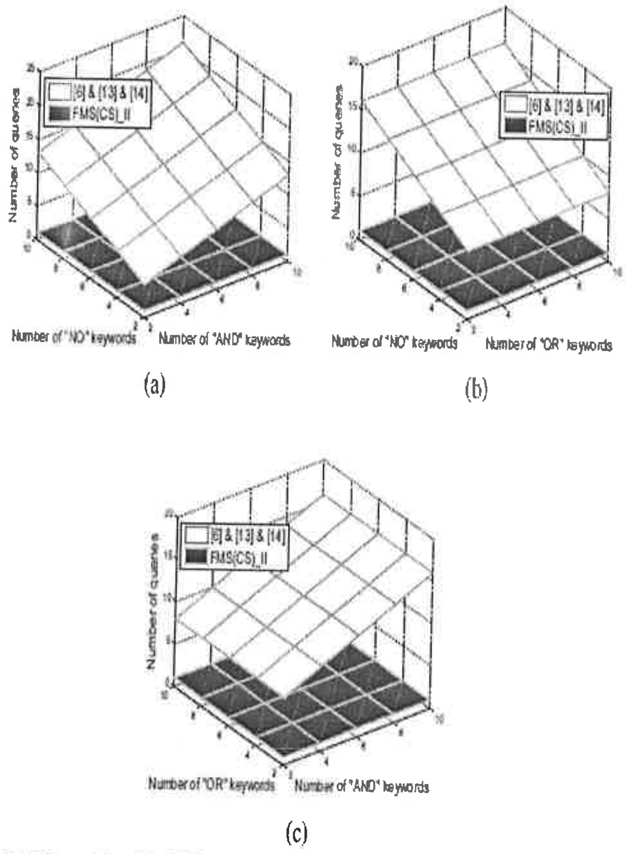


Fig.3. Time for Building Index. (a) Number of Queries for the Different Number of "AND" and "NO" Keywords with the Same Number of "OR" Keywords, L_1 = 5. (b) Number of Queries for the Different Number of "OR" and "NO" Keywords with the Same Number of "AND" Keywords, L_2 = 5. (C) Number of Queries for the Different Number of "AND" and "OR" Keywords with the Same Number of "NO" Keywords, L_3 = 5.

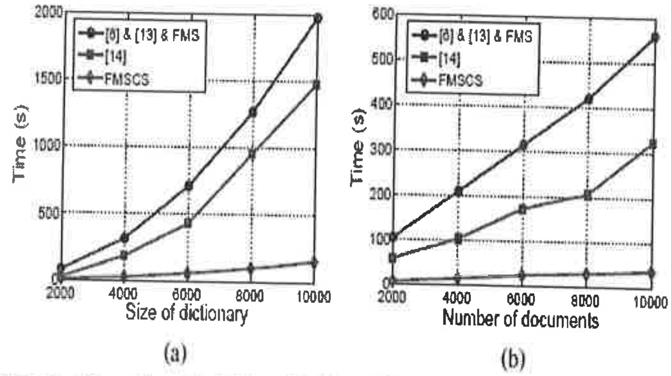


Fig.4. Time for Building Index. (a) For the Different Size of Dictionary with the Same Number of Documents, *N*=6000. (b) For the Different Number of Documents with the Same Size of Dictionary, |*W*| = 4000.

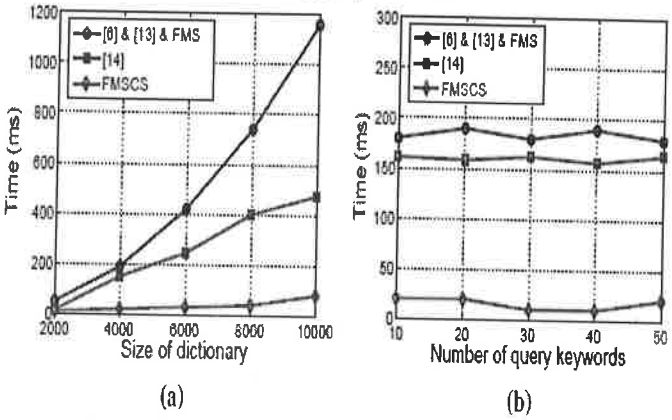


Fig.5. Time for Generating Trapdoor. (a) For the Different Size of Dictionary with the Same Number of Query Keywords, |$\widetilde{W}$|=20. (b) For the Different Number of Query Keywords with the Same Size of Dictionary, |*W*| = 4000.

In comparison, our FMS I and FMS II schemes should firstly generate a super-increasing sequence and a weight sequence, respectively. But actually, we can pre-select a corresponding sequence for each scheme, it can also achieve search process and privacy. Because even if the vectors are the same for multiple queries, the trapdoors will be not the same due to the security of kNN computation scheme. Therefore, the computation cost of [7] and all FMS schemes in *Trapdoor generating* phase are the same. As shown in Fig. 5, the time for generating trapdoor is dominated by the size of dictionary, instead of the number of query keywords. Hence, our FMSCS schemes are also very efficient in *Trapdoor generating* phase. Query. As [7] the FMS all adopt the secure kNN computation scheme, the time for query is the same. The computation overhead in *Query* phase, as shown in Fig. 6, is greatly affected by the size of dictionary and the number of documents, and almost has no relation to the number of query keywords. Further we can see, our FMSCS schemes significantly reduce the computation cost in *Query* phase. As needs to encrypt each dimension of index/trapdoor using full homomorphic encryption, its index/trapdoor size is enormous. Note that, in *Trapdoor generating* and *Query* phases, the computation overheads are not affected by the number of query keywords. Thus our FMS and FMSCS schemes are more efficient compared with some multiple keyword search schemes, as their cost is linear with the number of query keywords.

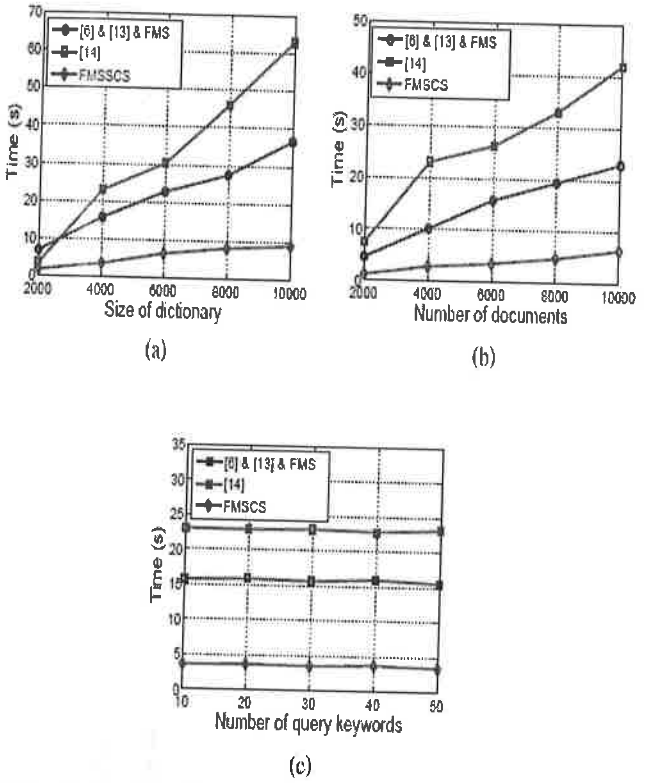


Fig.6. Time for Query. (a) For the Different Size of Dictionary with the Same Number of Documents and Number of Search Keywords, $N = 6000$; $|\widetilde{W}| = 20$. (b) for the Different Number of Documents with the Same Size of Dictionary and Number of Search Keywords, $|W| = 4000$; $|\widetilde{W}| = 20$. (c) For the Different Number of Search Keyword with the Same Size of Dictionary and Number of Documents, $N = 6000$; $|W| = 4000$.

2. Storage Overhead

As shown in Table 2, we provide a comparison of storage overhead among several schemes. Specifically, we evaluate the storage overhead from three parts: the data owner, the search user and the cloud server. According to Table 2, in the FMS, the FMSCS as well as schemes of [7], the storage overhead of the data owner is the same. In these schemes, the data owner preserves her secret key $K = (S, M_1, M_2)$ and symmetric key *sk* locally, where S is an $(m+1)$-dimensional vector, M_1 and M_2 are $(m+1) \times (m+1)$ invertible matrices. All elements in S, M_1 and M_2 are the float number. Since the size of a float number is 4 bytes, the size of K is $4\cdot(m+1) + 8\cdot(m + 1)2$ bytes. We assume that the size of *sk* is S_{sk} that is a constant. Thus, the total size of storage overhead is $4\cdot(m+1) +8\cdot(m + 1)2+S_{sk}$ bytes. However, in [14], the storage overhead of data owner is $\lambda^5/8$ bytes, where the λ is the secure parameter. The storage overhead is 4GB when we choose $\lambda = 128$, which is popular in a full homomorphic encryption scheme. However, the storage overhead of the FMS and the FMSCS are almost 763MB when we choose $m = 10000$, which is large enough for a search scheme. Therefore, the FMS and the FMSCS are more efficient than scheme in terms of the storage overhead of the data owner. As shown in Table II, a search user in the FMS, the FMSCS as well as the schemes of [7] preserves the secret key $K = (S, M_1, M_2)$ and the symmetric key *sk* locally.

Therefore, the total storage overhead is $4(m+1)+8(m + 1)^2+S_{sk}$ bytes. However, in, the storage overhead is $\lambda^5/8+ \lambda^2/8$ bytes. The storage overhead is 4GB when we choose $\lambda = 128$, which is popular in a full homomorphic encryption scheme. However, the storage overhead of the FMS and the FMSCS are almost 763MB when we choose $m = 10000$, which is large enough for a search scheme. Therefore, the FMS and the FMSCS are more efficient than scheme in terms of the storage overhead of the search user. The cloud server preserves the encrypted documents and the indexes. The size of encrypted documents in all schemes are the same, i.e., $N{\cdot}Ds$. For the indexes, in the FMS and schemes in [7], the storage overhead are $8 \cdot (m+1) \cdot N$ bytes. In the FMSCS, the storage overhead is $8\cdot''\cdot (m+1) \cdot N$ bytes, where $0 < \varepsilon < 1$. When $m = 1000$ and $N = 10000$ which are large enough for a search scheme, the storage overhead of indexes is about 132MB in the FMSCS. And in schemes of [7] as well as the FMS, the size of indexes is 760MB with the same conditions. In scheme the storage overhead of indexes is $N \cdot D_s + m \cdot N \cdot (\lambda/8)^5$ bytes, it is 4GB when we choose $\lambda = 128$, which is popular in a full homomorphic encryption scheme. Therefore, the FMS and the FMSCS are more efficient than scheme in terms of the storage overhead of the cloud server.

3. Communication Overhead

As shown in Table 3, we provide a comparison of communication overhead among several schemes. Specifically, we consider the communication overhead from three parts: the communication between the data owner and the cloud server (abbreviated as D-C), the communication between the search user and the cloud server (abbreviated as C-S) and the communication between the data owner and the search user (abbreviated as D-S). D-C. In the FMS as well as schemes of [7], the data owner needs to send information to cloud server in the form of $C_j \,||FID_j\, ||I_j$ ($j = 1; 2; \cdots ;N$), where the C_j represents the encrypted documents, FID_j represents the identity of the document and I_j represents the index. We assume that the average size of documents is D_s, thus the size of documents is $N \cdot D_s$. We assume the encrypted documents identity *FID* is a 10-byte string. Thus, the total size of the identity *FID* is $10 \cdot N$ bytes. The index $I_j = (paM_1, pbM_2)$ contains two (m+1)-dimensional vectors. Each dimension is a float number (the size of each float is 4 bytes). Thus, the total size of index is $8\cdot (m+1) \cdot N$ bytes. Therefore, the total size of communication overhead is $8\cdot (m+1) \cdot N+10{\cdot}N+N{\cdot}D_s$ bytes. In the FMSCS, the total size of communication overhead is $8\cdot''\cdot (m+1) \cdot N +10{\cdot}N+N{\cdot}D_s$ bytes. If we choose the ″ as 0.2, the size of index is $1.6 \cdot (m+1) \cdot N$ bytes, and the total size of communication of FMSCS is 1.6·

$(m+1)\cdot N+10\cdot N+D_s\cdot N$ bytes. However, in, the communication overhead is $N\cdot Ds+m\cdot N\cdot \lambda^5/8$ bytes, where λ is the secure parameter. If we choose $\lambda = 128$ which is popular in a full homomorphic encryption scheme and $m = 1000$ and $N = 10000$ which are large enough for a search scheme, the FMS and the FMSCS are more efficient than scheme in terms of the communication overhead of D-C.

TABLE II. Comparison of Storage Overhead (Bytes). (*M* Represents the Size of Dictionary; *N* Represents the Number of Documents; *D*$_s$ Represents the Average Size of Each Encrypted Document; Λ Represents the Secure Parameter; " Represents the Decrease Rate of Dictionary By Using Our Classified Sub-Dictionaries Technology; *S*$_{sk}$ Represents the Size of Symmetric Key.)

	[14]	[6], [13] and FMS	FMSCS
Data Owner	$\lambda^5/8$	$4\cdot(m+1)+8\cdot(m+1)^2+S_{sk}$	$4\cdot(m+1)+8\cdot(m+1)^2+S_{sk}$
Search User	$\lambda^5/8+\lambda^2/8$	$4\cdot(m+1)+8\cdot(m+1)^2+S_{sk}$	$4\cdot(m+1)+8\cdot(m+1)^2+S_{sk}$
Cloud Server	$N\cdot D_s+m\cdot N\cdot\lambda^5/8$	$N\cdot D_s+8\cdot(m+1)\cdot N$	$N\cdot D_s+8\cdot\varepsilon\cdot(m+1)\cdot N$

TABLE III. Comparison of Communication Overhead (Bytes). (*M* Represents the Size Of Dictionary; *N* Represents the Number of Documents; *D*$_s$ Represents the Average Size of Each Encrypted Document; *T* Represents the Number of Returned Documents; Λ Represents the Secure Parameter; " Represents the Decrease Rate of Dictionary by using our Classified Sub-Dictionaries Technology; *S*$_{sk}$ Represents the Size of Symmetric Key.)

	[14]	[6], [13] and FMS	FMSCS
D-C	$N\cdot D_s+m\cdot N\cdot\lambda^5/8$	$8\cdot(m+1)\cdot N+10\cdot N+N\cdot D_s$	$8\cdot\varepsilon\cdot(m+1)\cdot N+10\cdot N+N\cdot D_s$
C-S	$m\cdot\lambda^5/8+T\cdot D_s$	$8\cdot(m+1)+T\cdot D_s$	$8\cdot\varepsilon\cdot(m+1)+T\cdot D_s$
D-S	$\lambda^5/8+\lambda^2/8$	$4\cdot(m+1)+8\cdot(m+1)^2+S_{sk}$	$4\cdot(m+1)+8\cdot(m+1)^2+S_{sk}$

C-S. The C-S consists of two phases: *Query* and *Results returning*. In the *Query* phase, a search user in the FMS as well as the schemes in [7] sends the trapdoor to the cloud server in the form of $T_{\sim W} = (M^{-1}{}_1\, q_a, M^{-1}{}_2\, q_b)$, which contains two (m+1)-dimensional vectors. Thus, the communication overhead is $8\cdot(m+1)$ bytes. In the FMSCS, the communication overhead is $8\cdot\varepsilon\cdot(m+1)(0 <\varepsilon < 1)$ bytes. In the *Results returning* phase, the cloud server sends the corresponding result to the search user. The communication overhead of CS increases along with the number of returned documents at this point. We assume that the number of the returned documents is *T*, thus, the total communication overhead of cloud server to search user is $T\cdot Ds$ bytes. Therefore, the total communication overhead of C-S is $8\cdot m+T\cdot Ds$ bytes. In the FMS as well as the schemes in [7], the total communication overhead of C-S is $8\cdot\varepsilon\cdot(m+1)+T\cdot D_s$ bytes. In, the total communication overhead of C-S is $m\cdot\lambda^5/8+T\cdot Ds$ bytes. If we choose $\lambda = 128$ which is popular in a full homomorphic encryption scheme and $m = 1000$ and $N = 10000$ which are large enough for a search scheme, the FMS and the FMSCS are more efficient than scheme in terms of the communication overhead of C-S. D-S. From table 3, we can see that the communication overhead of the FMS, the FMSCS as well as schemes in [7] is the same. In the *Initialization* phase, the data owner sends the secret key $K = (S, M_1, M_2)$ and symmetric key *sk* to the search user, where *S* is an $(m+1)$- dimensional vector, M_1 and M_2 are $(m+1)\times(m+1)$ invertible matrices. Thus, the size of the secret key *K* is $4\cdot(m+1)+8\cdot(m+1)^2$ bytes. Therefore, the total size of communication overhead is $4\cdot(m+1)+8\cdot(m+1)^2+S_{sk}$ bytes, where the S_{sk} represents the size of symmetric key. However, the communication overhead of scheme is $\lambda^5/8+\lambda^2/8$ bytes. The communication overhead is 4GB when we choose $\lambda = 128$, which is popular in a full homomorphic encryption scheme. However, the communication overhead of the FMS and the FMSCS are almost 763MB when we choose $m = 10000$, which is large enough for a search scheme. Therefore, the FMS and the FMSCS are more efficient than scheme in terms of the communication overhead of D-S.

V. CONCLUSION

We have examined on the fine-grained multi keyword search (FMS) subject over encrypted cloud data, and future two FMS schemes. The FMS I includes both the significance scores and the partiality factors of keywords to augment more accurate search and enhanced users' experience, in that order. The FMS II realize secure and competent search with realistic functionality, i.e., "AND", "OR" and "NO" operations of keywords. In addition, we have planned the better schemes behind confidential sub-dictionaries (FMSCS) to advance competence. For the future work, we propose to add expand the application to reflect on the extensibility of the file set and the multi-user cloud environments. Towards this trend, we have made some beginning consequences on the extensibility and the multiuser cloud environments. Another remarkable topic is to increase the greatly scalable searchable encryption to enable able explore on large realistic databases.

VI. REFERENCES

[1]Hongwei Li, Member, IEEE, Yi Yang, Student Member, IEEE, Tom H. Luan, Member, IEEE, Xiaohui Liang, Student Member, IEEE, Liang Zhou, Member, IEEE, and Xuemin (Sherman) Shen, Fellow, IEEE, "Enabling Fine-grained Multi-keyword Search Supporting Classified Sub-dictionaries over Encrypted Cloud Data", IEEE Transactions on Dependable and Secure Computing, 2015.

[2]H. Liang, L. X. Cai, D. Huang, X. Shen, and D. Peng, "An SMDP based service model for inter domain resource allocation in mobile cloud networks," IEEE Transactions on Vehicular Technology, vol. 61, no. 5, pp. 2222–2232, 2012.

[3]M. M. Mahmoud and X. Shen, "A cloud-based scheme for protecting source-location privacy against hotspot-locating attack in wireless sensor networks," IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 10, pp. 1805–1818, 2012.

[4]Q. Shen, X. Liang, X. Shen, X. Lin, and H. Luo, "Exploiting redistributed clouds for e-health monitoring system with minimum service delay and privacy preservation," IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 2, pp. 430–439, 2014.

[5]T. Jung, X. Mao, X. Li, S.-J. Tang, W. Gong, and L. Zhang, "Privacy preserving data aggregation without secure

channel: multivariate polynomial evaluation," in Proceedings of INFOCOM. IEEE, 2013, pp. 2634–2642.

[6]Y. Yang, H. Li, W. Liu, H. Yang, and M. Wen, "Secure dynamic searchable symmetric encryption with constant document update cost," in Proceedings of GLOBCOM. IEEE, 2014, to appear.

[7]N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222–233, 2014.

[8]https://support.google.com/websearch/answer/173733?hl=en.

[9]D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proceedings of S&P. IEEE, 2000, pp. 44–55.

[10]R. Li, Z. Xu, W. Kang, K. C. Yow, and C.-Z. Xu, "Efficient multi-keyword ranked query over encrypted data in cloud computing," Future Generation Computer Systems, vol. 30, pp. 179–190, 2014.

[11]H. Li, D. Liu, Y. Dai, T. H. Luan, and X. Shen, "Enabling efficient multi-keyword ranked search over encrypted cloud data through blind storage," IEEE Transactions on Emerging Topics in Computing, 2014, DOI10.1109/ TETC.2014 .2371239.

[12]C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proceedings of ICDCS. IEEE, 2010, pp. 253–262.

Author's Profile:

G. Sandhya Rani Department of CSE, from Annamacharya Institute of Technology and Sciences, Hyderabad, TS, India.
E-mail: rockzsandhya@gmail.com.

Mr.Ramesh Babu Varugu received the Master of Technology degree in Information Technology from the Gurunanak Institute of Science And Technology-JNTUH, he received the Bachelor Of Engineering degree from Lakireddy Balireddy College of Engineering-JNTU-K. He is currently working as Associate Professor and a Head of the Department of CSE with Annamacharya institute of technology and sciences. His interest subjects are Operating Systems, Cloud computing.
E-mail: ramesh.vnl@gmail.com.

Mrs.V.Vedasahithi received the Master of Technology degree in Software Engineering from the St.Marys Group of Institutions-JNTUH, She received the Bachelor Of Engineering degree from K.L.UNIVERSITY. She is currently working as Assistant Professor with Annamacharya institute of technology and sciences. Her interest subjects are Adhoc networks, computer networks.
E-mail: sahithivellanki@gmail.com.

ijiTECH International Journal of Innovative Technologies

WWW.IJITECH.ORG

ISSN 2321-8665
Vol.04,Issue.13,
September-2016,
Pages:2317-2322

Improved Privacy-Preserving P2P Multimedia Distribution Based on Recombined Fingerprints

JAHNAVI PARVATHANENI[1], RAMESH BABU VARUGU[2], A.NAGA SRI[3]

[1]PG Scholar, Dept of CSE, Annamacharya Institute of Technology and Science, Hyderabad, TS, India,
E-mail: jahnavi.parvathanerii@gmail.com.

[2]Associate Professor& HOD, Dept of CSE, Annamacharya Institute of Technology and Science, Hyderabad, TS, India,
E-mail: ramesh.vnl@gmail.com.

[3]Assistant Professor, Dept of CSE, Annamacharya Institute of Technology and Science, Hyderabad, TS, India.

Abstract: Unknown unique mark has been proposed as a helpful answer for the lawful dissemination of sight and sound substance with copyright insurance while saving the security of purchasers, whose characters are just uncovered if there should arise an occurrence of unlawful re-appropriation. Notwithstanding, the vast majority of the current unknown fingerprinting conventions are unreasonable for two principle reasons: 1) the utilization of complex tedious conventions and/or homomorphic encryption of the substance, and 2) a unicast approach for conveyance that does not scale for an expansive number of purchasers. This paper originates from a past proposition of recombined fingerprints which conquers some of these disadvantages. In any case, the recombined unique finger impression approach requires a mind boggling chart hunt down deceiver following, which needs the investment of different purchasers, and genuine intermediaries in its P2P conveyance situation. This paper concentrates on evacuating these disservices bringing about a productive, adaptable, security safeguarding and P2P-based fingerprinting framework.

Keywords: P2P Content Distribution, Anonymous Finger Printing, Re-Distribution Tracing, Recombined Fingerprints.

I. INTRODUCTION

Fingerprinting digital contents is an attractive option to protect the rights of content authors and owners when contents are sold or otherwise distributed over the Internet. Basically, fingerprinting consists of embedding an imperceptible mark in the distributed content (which may be audio, pictures or video) to identify the content buyer. The embedded mark is different for each buyer, but the content should stay perceptually identical for all buyers. In case of illegal redistribution, the embedded mark will allow identifying the redistributors (against whom subsequent action might be taken). Most fingerprinting schemes can be classified in three types: symmetric, asymmetric and anonymous. In the first type, the merchant is the one who embeds the mark into the content; hence, the buyer cannot be formally accused of illegal redistribution, since the merchant also had access to the fingerprinted content and could be himself the re-distributor. In Asymmetric fingerprinting, the merchant does not have access to the fingerprinted copy, but he can recover the mark in case of illegal redistribution and thereby identify the malicious buyer. In anonymous fingerprinting, in addition to asymmetry, the buyer preserves her anonymity and hence she cannot be linked to the purchase of a specific content, except if she participates in an illegal redistribution. Fingerprinting schemes in the literature share the common feature of being centralized: in a way or another, the content owner/merchant has to be involved in the fingerprinting every time the content is sold to a certain buyer. Hence, distribution is basically unicast, which has the shortcoming that scalability is limited by the computing resources and bandwidth available at the content owner/ merchant.

This problem is further aggravated if one uses asymmetric or anonymous fingerprinting, which require more computation, communication and storage than the usual Symmetric fingerprinting. Multicast transmission of content, in which content is simultaneously transmitted to a group of receivers, is much more effective and bottleneck-free. Yet, multicast transmission does not allow sending different copies to each user, as required by fingerprinting schemes. So the question is: can we distribute fingerprinted content in a way that is more scalable than unicast transmission? P2P distribution systems allow answering the above question in the positive. In these systems, content receivers become senders to other users. This model can be viewed as an intermediate option between unicast and multicast. P2P distribution of all kinds of contents has become popular in recent years with the bandwidth increase of home communications are some example P2P protocols for private file exchange. It must be noted that P2P distribution is not limited to private users in home environments; some content providers are starting to facilitate P2P download of their products. Using a P2P system allows the merchant to establish only a small number of unicast connections with a set of M "seed" buyers who become new sources of content for other buyers.

The content can eventually reach a set of N buyers with M<< N. A. Contribution and plan of this paper We propose a P2P content distribution scheme (based on a specific P2P software) in which the merchant creates only a set of M seed copies of the content and sends them to M seed buyers. All subsequent copies are generated from the M seed copies. The copy obtained by a buyer is a combination of the copies supplied by her "parents" (sources). The fingerprint of each buyer is constructed as a binary sequence combining the sequences of her parents, in a way parallel to how DNA sequences of living beings are formed by combining the DNA sequences of their parents. The proposed scheme, which saves bandwidth and computation at the merchant, still allows tracking illegal redistributors but preserves the anonymity of honest buyers. The proposed method is thus inherently scalable compared to other systems in the literature, which require (non-scalable) unicast transmissions and rely on complex CPU-intensive and/or bandwidth consuming cryptographic protocols. The cryptographic protocols used in our approach reduce to the transmission of a few encrypted hashes with low computation and communication costs. In fact, the method proposed in this paper even avoids running the embedding algorithm for non-seed buyers and thus it outperforms the abovementioned methods.

II. DNA-INSPIRED FINGERPRINTS

In this section, we introduce a novel concept of automatic DNA-inspired binary fingerprints. The terms used in this paper are derived from those used in genetics to refer to DNA and heredity. The definitions of these terms in the context of this paper are introduced below. DNA sequence: in nature, DNA is a molecule consisting of an ordered set of nucleotides, where each nucleotide is one of the following four (smaller) molecules: adenine, cytosine, guanine and thymine, usually represented by their initial letters ("A", "C", "G" and "T"). Although the DNA molecule consists of two strings, the nucleotides are always paired A-T and C-G in the two different strings, meaning that the DNA structure is redundant. The DNA molecule is a double helix string in which each string is the complementary of the other one. Hence, one of the strings contains all the information required to build the other one. In this paper, a DNA-inspired fingerprint is constructed as a binary sequence and each bit can be considered as the counterpart of the nucleotides in real DNA sequences. Although each of the real DNA sequences' nucleotides can be thought of as a two-bit symbol (since there are four different nucleotides), the analogy can still be established using 1bit nucleotides. This is similar to what is done in genetic algorithms [9]. Gene: a segment of the DNA sequence which encodes a given protein –and thus has some impact in heredity and in the biological chemistry of the living being– is called a gene.

Similarly, a segment of the DNA-inspired fingerprint sequence is called a "gene" in this paper. Although real life genes have different sizes, the sizes of the genes in this paper are taken to be equal for simplicity and without loss of generality (variable sized genes might be used with no additional complexity in the proposed system). In addition, in nature, not all segments of the DNA sequence encode genes. Nevertheless, in this paper, all "nucleotides" (bits) do belong to one of the genes of the DNA-inspired fingerprint. Mating and heredity: in nature, the genes of an offspring are basically a combination of the genes of its parents (although some other processes such as mutation and crossover may produce fragments of DNA which are different in the offspring with respect to both its parents). Similarly, in this paper, when a buyer obtains a copy of a P2P-distributed content using some specific software, the DNA-inspired fingerprint of her copy will be a combination of the genes of the sources of the content (referred to as "parents" from the biological analogy). In this case, the number of parents for a buyer does not have to be exactly two as in the natural world. Hence, the mating process in the suggested fingerprinting scenario must be understood in a generalized sense, not limited to two parents. In this proposal, fingerprints can be considered as being "automatically generated" from the fingerprints of the parents. Despite this "automatic generation" of fingerprints, the constructed sequences are still valid for identification purposes, just like DNA traces can be used in criminal investigations to identify the suspect of an offence.

Mutation and crossover: different types of changes may occur in DNA molecules resulting in the modification of some of the nucleotides in the sequence. These changes may affect a single nucleotide or a full segment of the DNA sequence. Basically, crossover occurs when the two complementary DNA strings are recombined during DNA replication and mutation refers to different random-like errors during DNA replication. Mutation and crossover provide mechanisms which allow the DNA sequence of an offspring to include genes which are different from those of its parents. If it is allowed that a buyer can obtain her copy of the contents from only one parent (source), mutation (and/or crossover) shall be used to produce a different version of the fingerprint, as required in fingerprinting applications (since two different buyers must have different fingerprints). Note that, although the DNA inspired fingerprints are defined as a single bit stream, it is still possible to consider that a complementary sequence exists by using its negation. Crossovers can thus be simulated between a binary fingerprint and its negation. Although mutation and crossover provide a hypothetical mechanism to obtain a different fingerprint for parent and child in the single-parent case, still, the best and easiest strategy for a practical implementation of the scheme is to avoid this solution by enforcing at least two parents for each buyer.

The implementation presented in this paper uses neither mutation nor crossover since the system compels each buyer to obtain the content's fragments from at least two buyers. DNA relationship test: in real-world investigations, DNA relationship testing is often conducted to prove or disprove a blood relation between two or more individuals. Taking into account the mating and heredity processes and properties, blood relatives are known to share longer segments of their DNA sequences compared to those of non-relatives. The equivalent of this DNA relationship test in the proposed fingerprinting scheme is a function to compute the correlation between two binary strings. Since ancestors and descendants in the distributed fingerprinting scenario share several of their genes, a measurable correlation exists between their DNA inspired fingerprints.

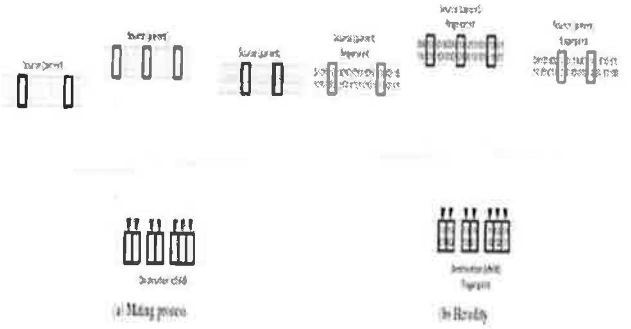


Fig.1.Upload/Download of the Content (Matching Process) and Automatic DNA-Inspired Fingerprint Construction (Heredity).

III. EXISTING AND PROPOSED SYSTEMS

A. Existing System

Most fingerprinting systems can be classified in three categories, namely symmetric, asymmetric and anonymous schemes. In symmetric schemes, the merchant is the one who embeds the fingerprint into the content and forwards the result to the buyer; hence, the buyer cannot be formally accused of illegal re-distribution, since the merchant also had access to the fingerprinted content and could be responsible for the re-distribution. In asymmetric fingerprinting, the merchant does not have access to the fingerprinted copy, but he can recover the fingerprint in case of illegal re-distribution and thereby identify the offending buyer. In anonymous fingerprinting, in addition to asymmetry, the buyer preserves her anonymity (privacy) and hence she cannot be linked to the purchase of a specific content, unless she participates in an illegal re-distribution.

B. Proposed System

The content is divided into several ordered fragments and each of them is embedded separately with a random binary sequence. The binary sequence for each fragment is called segment and the concatenation of all segments forms the whole fingerprint. The merchant distributes different copies to a reduced set of M seed buyers. The fingerprints of these buyers are such that their segments have low pair-wise correlations. The buyers other than the seed ones engage on P2P transfers of the content in such a way that each new buyer obtains fragments from at least two other buyers. The total number of buyers is N _ M. The communication between peer buyers is anonymous through an onion routing-like protocol using a proxy. The fingerprint of each new buyer is built as a recombination of the segments of its parents. Proxies know the pseudonyms of source and destination buyers and they have access to the symmetric keys used for encrypting the multimedia content. A transaction record is created by a transaction monitor to keep track of each transfer between peer buyers. These records do not contain the embedded fingerprints, but only an encrypted hash of them.

The fingerprints' hashes are encrypted in such a way that the private key of at least one parent is required for obtaining their clear text. The real identities of buyers are known only by the merchant. The transaction monitor records buyers' pseudonyms. In case of illegal re-distribution, a search is required through the distribution graph. The search starts from the seed buyers and is directed by a correlation function between the traced fingerprint and the fingerprints of the tested buyers. These tested buyers must co-operate with a tracing authority to compute the correlation between their fingerprint and the one extracted from the illegally re-distributed file. The fingerprints hashes recorded in the transaction monitor are enough to prevent buyers from cheating in this step. At each step of the traitor tracing protocol, the buyer with maximum correlation is chosen as the most likely ancestor of the illegal re-distributor. This criterion is mostly right, but some incorrect choices may occur during the search process, requiring the exhaustion of a sub graph and backtracking. The search ends when perfect correlation is found between the fingerprint of the tested buyer and that of the illegally re-distributed file. If a buyer refuses to take a correlation test, the hash recorded in the transaction monitor can be used as evidence against her.

1. Advantages of Proposed System

- This paper reviews the main features of the proposal suggested, highlights its main drawbacks, and suggests several significant improvements to achieve a more efficient and practical system, especially as traitor tracing is concerned, since it avoids the situations in which illegal redistributors cannot be traced with the proposal.
- Furthermore, better security properties against potentially malicious proxies are obtained.
- Although the system proposed in this paper uses public key encryption in the distribution and traitor tracing protocols, it must be taken into account that this encryption is only applied to short bit strings, such as the binary fingerprints and hashes, not to the content. The fragments of the content are encrypted using symmetric cryptography, which is much more efficient.

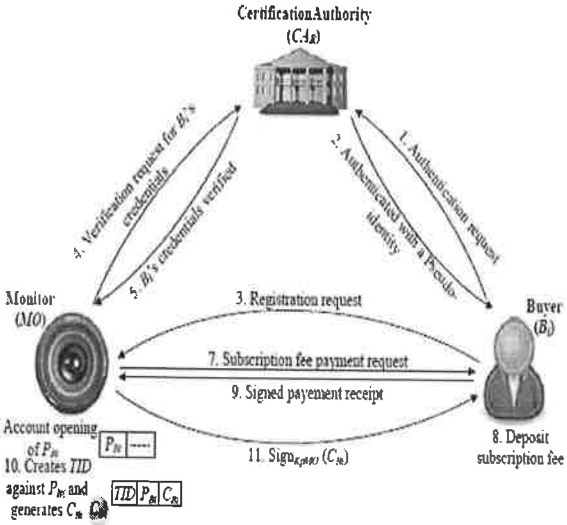

CertificationAuthority
(CA_R)
1. Authentication request
2. Authenticated with a Pseudo-identity
3. Registration request
4. Verification request for B_i's credentials
5. B_i's credentials verified
Monitor
(MO)
Buyer
(B_i)
7. Subscription fee payment request
9. Signed payement receipt
6. Account opening of P_{Bi}
P_{Bi}
----
10. Creates TID against P_{Bi} and generates C_{Bi}
TID
P_{Bi}
C_{Bi}
11. $\text{Sign}_{K_{p}MO}(C_{Bi})$
8. Deposit subscription fee

Fig.2. System Architecture.

IV. SECURITY AND PRIVACY ANALYSIS

This section analyzes the security and privacy properties of the proposed system according to the security model introduced. As detailed, attacks to the system may be classified as authentication/impersonation attacks, man-in the-middle attacks and protocol attacks. Authentication/ impersonation attacks should be overcome by using existing secure authentication protocols and are out of the scope of this paper. As man-in-the-middle attacks are concerned, there is no possibility of intercepting and decrypting the messages between a buyer and a proxy, since communications with the transaction monitor and the child buyer should also be attacked in order to obtain the session key used for encrypting the content. If the communication between the child buyer and the transaction monitor (Step 5 of Protocol 1) are strongly authenticated (e.g., using a Public Key Infrastructure), the possibility of a successful man-in the-middle attack can be neglected. The following sections deal with the security and privacy of the protocols proposed, first taking a formal approach and then with a description of more complex collusion attacks.

A. Formal Analysis of the Proposed Protocols

First of all, the security and privacy properties of Protocols 1 and 2 is analysed by means of two theorems (and their corresponding proofs).

1. Theorem 1: In Protocol 1, a malicious proxy trying to decrypt the fragments of the content would be detected.

Proof: If a malicious proxy tries to obtain the session key k by sending r to the transaction monitor there are two possibilities:

- If the child buyer has already retrieved k from the database by sending the handle r to the transaction monitor, the register containing k would be either blocked or removed. Note that the transaction monitor is assumed to be honest for the management of the symmetric keys.
- If the child buyer has not retrieved k from the transaction monitor, the proxy will obtain it, but the child buyer will find the corresponding register either blocked or removed. Then, the malicious behavior of the proxy can be reported to the authorities and the transaction monitor and the child buyer have enough information (such as pseudonyms and IP addresses) to identify the misbehaving proxy. Again, the assumption of honest behavior for the management of symmetric keys applies.

Hence, a malicious proxy trying to obtain k from r would be detected, since the register would be blocked either to the proxy or to the child buyer, raising an investigation. This completes the proof.

2. Theorem 2: By applying Protocol 2, an illegal re-distributor can be traced efficiently using a standard database search in the transaction monitor and it is not required to decrypt any of the fingerprints recorded by the transaction monitor. The output of the tracing protocol is the identity of at least one illegal re-distributor.

Proof: If no collusion occurs, the fingerprint f would be first extracted by the tracing authority, which is trusted. Then the tracing authority would compute $E^{c}_{gj}= E\ (g_j,\ K_c)$ for each segment (using the public key of the transaction monitor), and finally obtain E_f after grouping the segments in sets of m consecutive elements and encrypting these groups with its public key K_a. After that, the transaction monitor, which is also trusted for transaction database search, would output the pseudonym of the illegal re-distributor. The pseudonym can be linked to the real identity by the merchant, who provides also a signed document that associates the real identity and the pseudonym. This completes the proof.

In case of collusion of several buyers, the extracted fingerprint would not be a valid codeword of the anti-collusion code used in the scheme. Then, the system described would be used: the encrypted hash $E_{h,f}=E\ (h_f,\ K_c)$ would be searched instead of the encrypted fingerprint, where h_f denotes the hash obtained applying the hash function to the traced fingerprint f. Thus, Protocol 2 would be used with the hash of the fingerprint instead of the fingerprint itself. As described, with a large enough hash space, hash collisions would be almost negligible and a traitor would still be identified in the vast majority of the cases. The requirement that the transaction monitor is trusted and returns the pseudonym of the buyer associated with the traced fingerprint (and not a different pseudonym) can be relaxed if a signature of the encrypted sets of segments of the fingerprint is provided by the proxies. These signatures can be verified using the public keys of the proxies. In that case, both the signatures and the pseudonyms of the proxies shall also be included in the registers of the transaction database to facilitate the verification of these signatures when required.

B. Collusion Attacks on the Protocols

This section discusses possible collusion attacks on the proposed protocols.

1. Buyer Frame Proofness

As already discussed in, the merchant is not able to produce any buyer's fingerprint by random guess due to the numerical explosion of the fingerprint space, even with a reduced number of seed buyers on the other hand, the transaction monitor has access only to the hashes of the fingerprints (not the fingerprints themselves without the private key of the authority). Since the hash function is not invertible, it is not possible for the monitor (even in coalition with the merchant) to reconstruct any buyer's fingerprint. Possible collusions to disclose the specific fingerprint of an innocent buyer are the following:

- The tracing authority and the transaction monitor.
- All the proxies(for a transfer) and the transaction monitor.
- All the proxies (for a transfer) and the merchant.

In the first case, the authority and the transaction monitor may use their private keys to obtain the clear text of all the fingerprints. However, this possibility can be neglected since at least the authority must be trusted. In the second case, all the segments of the fingerprint could be decrypted using the private key of the transaction monitor, since the malicious proxies would not encrypt them with the public key of the authority. Also, the transaction monitor could collude with the proxies and use the session keys *k* to decrypt the fragments. Both possibilities would involve at least three malicious parties: all the proxies (two at least per each purchase) and the transaction monitor. In the third case, even if the transaction monitor does not provide her private key, a brute force attack segment by segment would be possible to reconstruct a buyer's fingerprint, because the number of different segments is small for each fragment (equal to *M*). Again, at least three malicious parties would be required: two (or more) proxies plus the merchant. Hence, the minimum coalition required to frame an innocent buyer is formed by three malicious parties (or two if one of them is the authority). Note that a coalition of the transaction monitor and the merchant is not enough to obtain the clear text of any fingerprint. As the proxies encrypt a set of m consecutive segments, and there are *M* possible values for each segment, the total number of combinations per set of consecutive segments is M^m. This avoids a brute force attack if m is reasonably large. For example, if $M = 10$ and $m = 32$, there would be 10^{32} possible combinations for each set of consecutive segments, what would be enough for security against a brute force attack if the segments were encrypted one by one (or grouped with a small value of *m*), the system would be vulnerable against a brute force attack for a collusion of the merchant and the transaction monitor.

2. Copyright Protection

In order to ensure copyright protection, it is essential that the fingerprint embedded in each buyer's copy of the content and its encrypted version recorded by the transaction monitor is identical. If there is a way to cheat in the recorded fingerprint, the corresponding buyer would be able to re-distribute her copy illegally without any chance of being detected. As already remarked in, the content fragments are signed by the merchant from origin. The same approach can be used here for each encrypted segment of the fingerprint, making it impossible for a proxy to cheat about the fingerprint. The authority and the merchant could verify randomly, with some probability, the signatures of the set of contiguous segments reported by a proxy. If the signature was not verified, the proxy would be accused of forgery. Note that the fingerprints would still be protected since 1) only some sets of contiguous segments would be verified (not the whole fingerprint) and 2) those segments would still be encrypted with the transaction monitor's public key. However, a proxy may still try to get alternative fragments for the same position of the content by requesting them from different parents. That possibility would allow the proxy to cheat about the true fingerprint of the child buyer, since several correctly signed fragments would be available for him for the same content. This behavior can be avoided in several ways. For example, temporary records can be created in the transaction monitor by the parents to detect if a proxy tries to obtain two alternative fragments for the same content.

3. Buyers' Privacy

The identity of a buyer who has purchased a specific content could be revealed by a coalition of two parties: one of the proxies chosen by the buyer and the merchant (who can link her pseudonym to a real identity) or, similarly, the transaction monitor and the merchant. Better privacy could be achieved if, for example, the pseudonyms were encrypted by the proxies using the public key of the tracing authority. In that case, a coalition of the merchant and the transaction monitor would not be enough to break a buyer's privacy, but a coalition of a proxy and the merchant would still be enough. However, the merchant should not be interested, in principle, to break her client's privacy, since privacy would be one of the clear advantages of the proposed distribution system. Another threat to privacy is the fact that all anonymous communications between each child and each parent occur through a unique proxy. This means that this proxy has access to different pseudonyms (the parents' and the child's). This can be easily circumvented if more proxies are used in Protocol 1 between child and parent. With two proxies, each of them would know only the pseudonym of one of the parties (although they could still collide). With three or more proxies, only two of them would have access to different pseudonyms (either the parents' or the child's). Of course, increasing the number of proxies in each transfer would affect the efficiency of the system, since more communication burden would be required.

V. CONCLUSION

A DNA-inspired fingerprinting scheme designed for P2P content distribution is presented. The proposed scheme allows the merchant to trace traitors who redistribute the content illegally. The merchant knows at most the fingerprinted copies of the seed buyers, but not the fingerprinted copies of non-seed buyers (the vast majority). Hence, the merchant does not know the identities of non-seed buyers. Whenever a traitor needs to be traced; only a small fraction of honest users must cooperate by providing their fingerprinted copies (quasi-privacy) collusion resistance against dishonest buyers trying to create a forged copy without any of their fingerprints is also discussed. Finally, buyer frame proofness is guaranteed since a malicious merchant does not have access to the fingerprinted copies of non-seed nodes. Thus, he will not be able to frame an honest buyer since random guess is not an option to construct a valid fingerprint (due to combinatorial explosion). Future research will involve designing backtrack-free protocols for traitor tracing in such a way that the fraction of honest buyers who must co-operate in case of an illegal redistribution is reduced. The security analysis of the proposed scheme against malicious proxies, who may even collude with other parties, is also left for the future research.

VI. REFERENCES

[1]David Megias, Member, IEEE, "Improved Privacy-Preserving P2P Multimedia Distribution Based on Recombined Fingerprints", IEEE transactions on dependable and secure computing, vol. 12, no. 2, march/april, 2015.

[2]D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," in Proc. 15th Ann. Int. Cryptology Conf. Adv. Cryptology, 1995, pp. 452–465.

[3]Y. Bo, L. Piyuan, and Z. Wenzheng, "An efficient anonymous fingerprinting protocol," in Proc. Int. Conf. Comput. Intell. Security, 2007, pp. 824–832.

[4]J. Camenisch, "Efficient anonymous fingerprinting with group signatures," in Proc. 6th Int. Conf. Theory Appl. Cryptology Inf. Security: Adv. Cryptology, 2000, pp. 415–428.

[5]C.-C. Chang, H.-C. Tsai, and Y.-P. Hsieh, "An efficient and fair buyer-seller fingerprinting scheme for large scale networks," Comput. Security, vol. 29, pp. 269–277, Mar. 2010.

[6]D. L. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," Commun. ACM, vol. 24, pp. 84–90, Feb. 1981.

[7]I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker, Digital Watermarking and Steganography. Burlington, MA, USA: Morgan Kaufmann, 2008.

[8]J. Domingo-Ferrer and D. Megıas, "Distributed multicast of fingerprinted content based on a rational peer-to-peer community," Comput. Commun., vol. 36, pp. 542–550, Mar. 2013.

[9]M. Fallahpour and D. Megıas, "Secure logarithmic audio watermarking scheme based on the human auditory system," Multimedia Syst., vol. 20, pp. 155–164, 2014.

[10]S. Katzenbeisser, A. Lemma, M. Celik, M. van der Veen, and M. Maas, "A buyer-seller watermarking protocol based on secure embedding," IEEE Trans. Inf. Forensics Security, vol. 3, no. 4, pp. 783–786, Dec. 2008.

[11]M. Kuribayashi, "On the implementation of spread spectrum fingerprinting in asymmetric cryptographic protocol," EURASIP J. Inf. Security, vol. 2010, pp. 1:1–1:11, Jan. 2010.

[12]C.-L. Lei, P.-L. Yu, P.-L. Tsai, and M.-H. Chan, "An efficient and anonymous buyer-seller watermarking protocol," IEEE Trans. Image Process., vol. 13, no. 12, pp. 1618–1626, Dec. 2004.

Author's Profile:

Jahnavi Parvathaneni Department of CSE, From Annamacharya Institute of Technology and Science, Hyderabad, TS, India. E-mail: jahnavi.parvathaneni@gmail.com.

Mr.Ramesh Babu Varugu, received the Master of Technology degree in Information Technology from the Gurunanak Institute of Science And Technology-JNTUH, he received the Bachelor of Technology degree from Lakireddy Balireddy College of Engineering,JNTUK. He is currently working as Associate Professor and a Head of the Department of CSE with Annamacharya Institute of Technology And Sciences, hyderabad. His interest subjects are operating Systems,Cloud Cimputing and etc.
E-mail: ramesh.vnl@gmail.com.

Mrs A.Naga Sri, received the Master of Technology degree in Computer science and Engineering from Acharya Nagarjuna University, she received the Bachelor of Engineering degree from Avanthi Engineering College-JNTUH. She is currently working as Assistant Professor of CSE with Annamachraya Institute Of Technology And Sciences, Hyderabad. Her interest subjects are CLDS, Computer Networks, Compiler Design.

ijiTECH International Journal of Innovative Technologies
WWW.IJITECH.ORG
ISSN 2321-8665
Vol.04,Issue.13,
September-2016,
Pages:2330-2334

A Secure Anti-Collusion Data Sharing Scheme for Dynamic Groups in the Cloud

ANAGANTI SUDHA[1], MR.RAMESH BABU VARUGU[2], A.NAGA SRI[3]
[1]PG Scholar, Dept of CSE, Annamacharya Institute of Technology and Science, Hyderabad, TS, India,
E-mail: sudha.anaganthi@gmail.com.
[2]Associate Professor & HOD, Dept of CSE, Annamacharya Institute of Technology and Science, Hyderabad, TS, India,
E-mail: ramesh.vnl@gmail.com.
[3]Assistant Professor, Dept of CSE, Annamacharya Institute of Technology and Science, Hyderabad, TS, India.

Abstract: The Benefited from Cloud Computing, clients can achieve a flourishing and moderate methodology for information sharing among gathering individuals in the cloud with the characters of low upkeep and little administration cost. Then, security certifications to the sharing information records will be given since they are outsourced. Horribly, due to the never-ending change of the enrolment, sharing information while giving protection saving is still a testing issue, particularly for an untrusted cloud because of the agreement attack. In addition, for existing plans, the security of key dispersion depends on the safe communication channel, then again, to have such channel is a solid feeling and is difficult for practice. In this paper, we propose a safe information sharing plan for element individuals. Firstly, we propose a safe route for key dispersion with no safe correspondence channels, and the clients can safely acquire their private keys from gathering administrator. Besides, the plan can accomplish fine-grained access control, any client in the gathering can utilize the source in the cloud and refused clients can't get to the cloud again after they are rejected. Thirdly, we can protect the plan from trickery attack, which implies that rejected clients can't get the first information record regardless of the possibility that they scheme with the untrusted cloud. In this methodology, by utilizing polynomial capacity, we can achieve a protected client denial plan. At long last, our plan can bring about fine productivity, which implies past clients need not to overhaul their private keys for the circumstance either another client joins in the gathering or a client is give up from the gathering.

Keywords: Access Control, Privacy-Preserving, Cloud Computing, Key Distribution.

I. INTRODUCTION

Cloud computing, with the characteristics of intrinsic data sharing and low maintenance, provides a better utilization of resources. In cloud computing, cloud service providers offer an abstraction of infinite storage space for clients to host data. It can help clients reduce their financial overhead of data managements by migrating the local managements system into cloud servers. However, security concerns become the main constraint as we now outsource the storage of data, which is possibly sensitive, to cloud providers. To preserve data privacy, a common approach is to encrypt data files before the clients upload the encrypted data into the cloud. Unfortunately, it is difficult to design a secure and efficient data sharing scheme, especially for dynamic groups in the cloud. A cryptographic storage system that enables secure data sharing on untrustworthy servers based on the techniques that dividing files into file groups and encrypting each file group with a file-block key. However, the file-block keys need to be updated and distributed for a user revocation; therefore, the system had a heavy key distribution overhead. However, the complexities of user participation and revocation in these schemes are linearly increasing with the number of data owners and the revoked users. The techniques of key policy attribute-based encryption, proxy re-encryption and lazy re-encryption to achieve fine-grained data access control without disclosing data contents.

However, the single-owner manner may hinder the implementation of applications, where any member in the group can use the cloud service to store and share data files with others. However, the scheme will easily suffer from the collusion attack by the revoked user and the cloud. The revoked user can use his private key to decrypt the encrypted data file and get the secret data after his revocation by conspiring with the cloud. In the phase of file access, first of all, the revoked user sends his request to the cloud, and then the cloud responds the corresponding encrypted data file and revocation list to the revoked user without verifications. Next, the revoked user can compute the decryption key with the help of the attack algorithm. Finally, this attack can lead to the revoked users getting the sharing data and disclosing other secrets of legitimate members. Unfortunately, the secure way for sharing the personal permanent portable secret between the user and the server is not supported and the private key will be disclosed once the personal permanent portable secret is obtained by the attackers. In this paper, we propose a secure data sharing scheme, which can achieve secure key distribution and data sharing for dynamic group. The main contributions of our scheme include:

- We provide a secure way for key distribution without any secure communication channels. The users can securely obtain their private keys from group manager without any Certificate Authorities due to the verification for the public key of the user.

- Our scheme can achieve fine-grained access control, with the help of the group user list, any user in the group can use the source in the cloud and revoked users cannot access the cloud again after they are revoked.
- We propose a secure data sharing scheme which can be protected from collusion attack. The revoked users can not be able to get the original data files once they are revoked even if they conspire with the untrusted cloud. Our scheme can achieve secure user revocation with the help of polynomial function.
- Our scheme is able to support dynamic groups efficiently, when a new user joins in the group or a user is revoked from the group, the private keys of the other users do not need to be recomputed and updated.
- We provide security analysis to prove the security of our scheme. In addition, we also perform simulations to demonstrate the efficiency of our scheme.

II.EXISTING AND PROPOSED SYSTEMS

A. Existing System

Kallahalla et al presented a cryptographic storage system that enables secure data sharing on untrustworthy servers based on the techniques that dividing files into file groups and encrypting each file group with a file-block key. Yu et al exploited and combined techniques of key policy attribute-based encryption, proxy re-encryption and lazy re-encryption to achieve fine-grained data access control without disclosing data contents.

B. Proposed System

In this paper, we propose a secure data sharing scheme, which can achieve secure key distribution and data sharing for dynamic group. We provide a secure way for key distribution without any secure communication channels. The users can securely obtain their private keys from group manager without any Certificate Authorities due to the verification for the public key of the user. Our scheme can achieve fine-grained access control, with the help of the group user list, any user in the group can use the source in the cloud and revoked users cannot access the cloud again after they are revoked. We propose a secure data sharing scheme which can be protected from collusion attack. The revoked users can not be able to get the original data files once they are revoked even if they conspire with the untrusted cloud. Our scheme can achieve secure user revocation with the help of polynomial function. Our scheme is able to support dynamic groups efficiently, when a new user joins in the group or a user is revoked from the group, the private keys of the other users do not need to be recomputed and updated. We provide security analysis to prove the security of our scheme.

Advantages of Proposed System:

- The computation cost is irrelevant to the number of revoked users in RBAC scheme. The reason is that no matter how many users are revoked, the operations for members to decrypt the data files almost remain the same.
- The cost is irrelevant to the number of the revoked users. The reason is that the computation cost of the cloud for file upload in our scheme consists of two verifications for signature, which is irrelevant to the number of the revoked users. The reason for the small computation cost of the cloud in the phase of file upload in RBAC scheme is that the verifications between communication entities are not concerned in this scheme.
- In our scheme, the users can securely obtain their private keys from group manager Certificate Authorities and secure communication channels. Also, our scheme is able to support dynamic groups efficiently, when a new user joins in the group or a user is revoked from the group, the private keys of the other users do not need to be recomputed and updated.

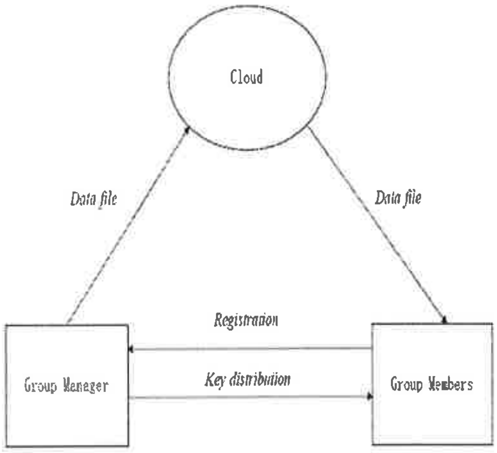


Fig.1. System Architecture.

III. SYSTEM MODEL

A. Threat Model

In this paper, we propose our plan taking into account the Dolev-Yao model, in which the attacker can catch, capture and combination any message at the correspondence channels with the Dolev-Yao model, the best way to protect the data from attack.

B. System Model

Here the proposed model is illustrated in Fig.1; the system model consists of three different entities: the cloud, a group manager and a large number of group members. The cloud, sustaining by the cloud service providers, provides storage space for hosting data files in a pay-as-you-go manner. on the other hand, the cloud is untrusted since the cloud service providers are easily to become untrusted. Therefore, the cloud will try to learn the content of the stored data. Group manager will obtain charge of system parameters generation, user registration, also, client repudiation. Bunch individuals (clients) are an arrangement of sign up clients that will store their own particular information into the cloud and impart them to others. In the plan, the gathering enrollment is powerfully changed, because of the new client call-up and client denial.

C. Design Goals

We depict the principle plan objectives of the proposed plan including key circulation, information secrecy, access control and effectiveness as takes after:

1. Key Distribution: The prerequisite of key transportation is that clients can safely get their private keys from the gathering director with no Certificate Authorities. In other existing plans, this purpose is skilful by expecting that the communication channel is secure, on the other hand, in our plan, we can accomplish it without this solid thought.

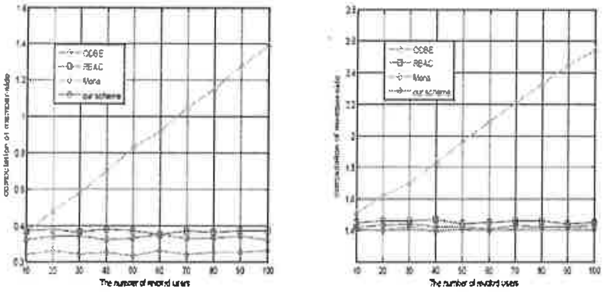
2. Access Control: First, collect individuals can make use of the cloud asset for information stockpiling and information sharing. Second, unapproved clients can't get to the cloud asset whenever, and disavowed clients will be unfitted for utilizing the cloud asset again once they are renounced.

3. Information Classification: Data secrecy requires that unapproved clients including the cloud are unequipped for taking in the substance of the put away information. To keep up the accessibility of information secrecy for element gatherings is still an essential and testing issue. In particular, renounced clients can't unscramble the put away information document after the denial. Effectiveness: Any gathering part can store and impart information records to others in the gathering by the cloud. Client repudiation can be accomplished without including the others, which implies that the remaining clients don't have to overhaul their private keys.

IV. PERFORMANCE EVALUATION

We make the performance simulation with NS2 and compare with Mona in [11] and the original dynamic broadcast encryption (ODBE) scheme. Without loss of generality, we set p =160 and the elements in G_1 and G_2 to be 161 and 1,024 bits, respectively. In addition, we assume the size of the data identity is 2^{16} bits, which yield a group capacity of data files. Similarly, the size of user and group identity are also set 16 bits. Both group members and group managers processes are conducted on a laptop with Core 2 T5800 2.0 GHz, DDR2 800 2G, Ubuntu 12.04 X86. The cloud process is implemented on a laptop with Core i7-3630 2.4 GHz, DDR3 1600 8G, Ubuntu 12.04 X64. We select an elliptic curve with 160 bits group order.

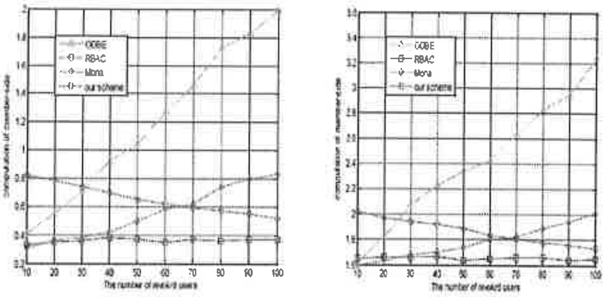
A. Member Computation Cost



(a) Generating a 10 MB file (b) Generating a 100 MB file

Fig.2. Comparison on Computation Cost of Members for FileUpload among ODBE, RBAC, Mona and Our Scheme.

As illustrated in Fig2, we list the comparison on computation cost of members for file upload among ODBE, RBAC, Mona and our scheme. It is obviously observed that the computation cost for members in our scheme is irrelevant to the number of revoked users. The reason is that in our scheme, we move the operation of user revocation to the group manager so that the legal clients can encrypt the data files alone without involving information of other clients, including both legal and revoked clients. On the contrary, the computation cost increases with the number of revoked users in ODBE. The reason is that several operations including point multiplications and exponentiations have to be performed by clients to compute the parameters in ODBE.



(a) Downloading a 10 MB file (b) Downloading a 100 MB file

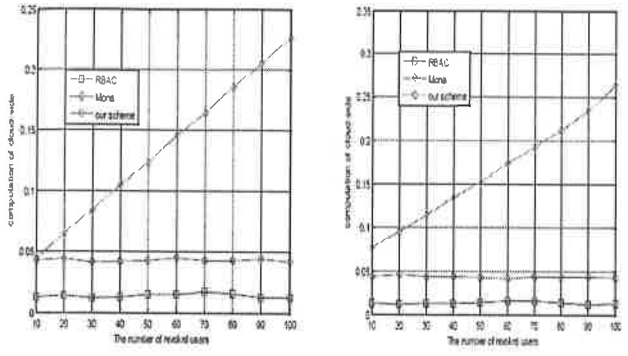
Fig.3. Comparison on Computation Cost of Members for File Download among ODBE, RBAC, Mona and Our Scheme

The computation cost of members for file download operations with the size of 10 and 100Mbytes are illustrated in Fig.3. The computation cost is irrelevant to the number of revoked users in RBAC scheme. The reason is that no matter how many users are revoked, the operations for members to decrypt the data files almost remain the same. The computation cost in Mona increases with the number of revoked users, because the users need to perform computing for revocation verification and check whether the data owner is a revoked user. Besides the above operations, more parameters need to be computed by members in ODBE. On the contrary, the computation cost decreases with the number of revoked users in our scheme because of the computation for the recovery of the secret parameter decreases with the number of revoked users.

B. Cloud Computation Cost

As illustrated in Fig.4, we list the comparison on computation cost of the cloud for file upload between Mona and our scheme. In general, it can be obviously seen that both the computation costs of the cloud in two schemes are acceptable. In detail, the cost in Mona increases with the number of revoked users, as the revocation verification cost increases. However, in our scheme, the cost is irrelevant to the number of the revoked users. The reason is that the computation cost of the cloud for file upload in our scheme consists of two verifications for signature, which is irrelevant to the number of the revoked users. The reason for the small computation cost of the cloud in the phase of file upload in

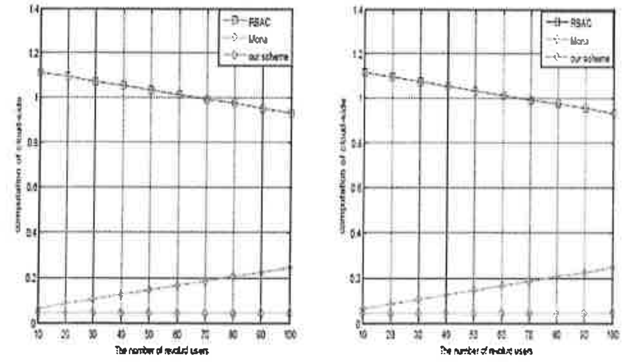
RBAC scheme is that the verifications between communication entities are not concerned in this scheme.



(a) Uploading a 10 MB file (b) Uploading a 100 MB file

Fig.4. Comparison on Computation Cost of Members for File Upload among RBAC, Mona and Our Scheme

The computation cost of the cloud for file download operations with the size of 10 and 100Mbytes are illustrated in Fig.5. Similar to the operation of file upload, the computation cost of the cloud is mainly determined by the revocation verification operation. Therefore, the cost increases with the number of revoked users. However, in our scheme, the cloud just simply verifies the signature. Therefore, the computation cost of the cloud for file download is irrelevant to the number of the revoked users. The reason for the high computation cost of the cloud in RBAC scheme is that the cloud performs some algorithm operations to help the user to decrypt data files. In addition, it can be seen that in these schemes, the computation cost is independent with the size of the file, since both the signature in Mona and the encrypted message in our scheme are irrelevant to the size of the requested file and the operations of cloud for decryption in RBAC scheme is also irrelevant to the size of the encrypted data files.



(a) Downloading a 10 MB file (b) Downloading a 100 MB file

Fig.5. Comparison on Computation Cost of the Cloud for File Download among RBAC, Mona and Our Scheme

V. CONCLUSION

In this paper, we outline a protected against agreement information sharing plan for element bunches in the cloud. In our plan, the clients can safely acquire their private keys from gathering director Certificate Authorities and secure correspondence channels. Likewise, our plan can bolster dynamic gatherings proficiently, when another client joins in the gathering or a client is denied from the gathering, the private keys of alternate clients don't should be recomputed and redesigned. In addition, our plan can accomplish secure client repudiation, the disavowed clients can not have the capacity to get the first information records once they are denied regardless of the possibility that they plot with the untrusted cloud.

VI. REFERENCES

[1]Zhongma Zhu, Rui Jiang, "A Secure Anti-Collusion Data Sharing Scheme for Dynamic Groups in the Cloud", IEEE Transactions on Parallel and Distributed Systems, 2015.

[2]M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. "A View of Cloud computing," Comm. ACM, vol. 53, no. 4, pp. 50-58, Apr.2010.

[3]S. Kamara and K. Lauter, "Cryptographic Cloud Storage," Proc. Int'l Conf. Financial Cryptography and Data Security (FC), pp.136- 149, Jan. 2010.

[4]M. Kallahalla, E. Riedel, R. Swaminathan, Q. Wang, and K. Fu, "Plutus: Scalable Secure File Sharing on Untrusted Storage," Proc. USENIX Conf. File and Storage Technologies, pp. 29-42, 2003.

[5]E. Goh, H. Shacham, N. Modadugu, and D. Boneh, "Sirius: Securing Remote Untrusted Storage," Proc. Network and Distributed Systems Security Symp. (NDSS), pp. 131-145, 2003.

[6]G. Ateniese, K. Fu, M. Green, and S. Hohenberger, "Improved Proxy Re-Encryption Schemes with Applications to Secure Distributed Storage," Proc. Network and Distributed Systems Security Symp. (NDSS), pp. 29-43, 2005.

[7]Shucheng Yu, Cong Wang, Kui Ren, and Weijing Lou, "Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing," Proc. ACM Symp. Information, Computer and Comm. Security, pp. 282-292, 2010.

[8]V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-Based Encryption for Fine-Grained Access Control of Encrypted Data," Proc. ACM Conf. Computer and Comm. Security (CCS), pp. 89-98, 2006

[9]R. Lu, X. Lin, X. Liang, and X. Shen, "Secure Provenance: The Essential of Bread and Butter of Data Forensics in Cloud Computing," Proc. ACM Symp. Information, Computer and Comm. Security, pp. 282-292, 2010.

[10]B.Waters,"Ciphertext-Policy Attribute-Based Encryption: An Expressive, Efficient, and Provably Secure Realization," Proc. Int'l Conf.http://eprint.iacr .org/ 2008 /290 .pdf, 2008

[11]Xuefeng Liu, Yuqing Zhang, Boyang Wang, and Jingbo Yang, "Mona: Secure Multi-Owner Data Sharing for Dynamic Groups in the Cloud," IEEE Transactions on Parallel and Distributed Systems, vol. 24, no. 6, pp. 1182-1191, June 2013.

[12]D. Boneh, X. Boyen, and E. Goh, "Hierarchical Identity Based Encryption with Constant Size Ciphertext," Proc. Ann. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT), pp. 440-456, 2005.

Author's Profile:

Anaganti Sudha, Department of CSE, from Annamacharya Institute of Technology and Science, Hyderabad, TS, India.
E-mail: sudha.anaganti@gmail.com.

Mr.Ramesh Babu Varugu, received the Master of Technology degree in information technology from the Gurunanak Institute of Science and Technology-JNTUH, he received the bachelor of technology degree from Lakireddy Balireddy College of Engineering, JNTUK. He is currently working as Associate Professor and a head of the Department of CSE with Annamacharya Institute of Technology and Sciences, Hyderabad. His interest subjects are Operating Systems, Cloud Computing etc. E-mail: ramesh.vnl@gmail.com.

Mrs A.Naga Sri, received the Master of Technology degree in Computer science and Engineering from Acharya Nagarjuna University, she received the Bachelor of Engineering Degree from Avanthi Engineering College-JNTUH. She is currently working as Assistant Professor of CSE with Annamachraya Institute of Technology and Sciences, Hyderabad. Her interest subjects are CLDS, Computer Networks, Compiler Design.

IJITECH International Journal of Innovative Technologies
WWW.IJITECH.ORG
ISSN 2321-8665
Vol.04,Issue.13,
September-2016,
Pages:2335-2340

CLOUD ARMOR: Supporting Reputation-Based Trust Management for Cloud Services

DASARI SWAPNA[1], J. RAMESH BABU VARUGU[2], B.RAVINDER REDDY[3]
[1]PG Scholar, Dept of CSE, Annamacharya Institute of Technology and Science, Hyderabad, TS, India,
E-mail: dasariswapna912@gmail.com.
[2]Associate Professor & HOD, Dept of CSE, Annamacharya Institute of Technology and Science, Hyderabad, TS, India,
E-mail: hodeceact@gmail.com.
[3]Assistant Professor, Dept of CSE, Annamacharya Institute of Technology and Science, Hyderabad, TS, India,
E-mail: p.satish605@gmail.com.

Abstract: In cloud computing growth, the management of trust element is most challenging issue. Cloud computing has produce high challenges in security and privacy by the changing of environments. Trust is one of the most concerned obstacles for the adoption and growth of cloud computing. Although several solutions have been proposed recently in managing trust feedbacks in cloud environments, how to determine the credibility of trust feedbacks is mostly neglected. In this project the system proposed a Cloud Armor, a reputation-based trust management framework that provides a set of functionalities to deliver Trust as a Service (TaaS). "Trust as a Service" (TaaS) framework to improve ways on trust management in cloud environments. The approaches have been validated by the prototype system and experimental results.

Keywords: Cloud Computing, Credibility, Credentials, Trust Management, Reputation, Security, Privacy, Availability.

I. INTRODUCTION

The highly dynamic, distributed, and nontransparent nature of cloud services make the trust management in cloud environments a significant challenge.According to researchers at Berkeley, trust and security is ranked one of the top 10 obstacles for the adoption of cloud computing. Indeed, Service-Level Agreements (SLAs) alone are inadequate to establish trust between cloud consumers and providers because of its unclear and inconsistent clauses. Consumers' feedback is a good source to assess the overall trustworthiness of cloud services. Several researchers have recognized the significance of trust management and proposed solutions to access and manage trust based on feedbacks collected from participants. In reality, it is not unusual that a cloud service experiences malicious behaviors (e.g., collusion or Sybil attacks) from its users. This paper focuses on improving trust management in cloud environments by proposing novel ways to ensure the credibility of trust feedbacks. In particular we distinguish the following key issues of the trust management in cloud environments: Consumers' Privacy. The adoption of cloud computing raise privacy concerns .Consumers can have dynamic interactions with cloud providers, which may involve sensitive information. There are several cases of privacy breaches such as leaks of sensitive information (e.g., date of birth and address) or behavioral information (e.g., with whom the consumer interacted, the kind of cloud services the consumer showed interest, etc.).

Undoubtedly, services which involve consumers data (e.g., interaction histories) should preserve their privacy. Cloud Services Protection. It is not unusual that a cloud service experiences attacks from its users. Attackers can disadvantage a cloud service by giving multiple misleading feedbacks (i.e., collusion attacks) or by creating several accounts (i.e., Sybil attacks). Indeed, the detection of such malicious behaviors poses several challenges. Firstly, new users join the cloud environment and old users leave around the clock. This consumer dynamism makes the detection of malicious behaviors (e.g., feedback collusion) a significant challenge. Secondly, users may have multiple accounts for a particular cloud service, which makes it difficult to detect Sybil attacks. Finally, it is difficult to predict when malicious behaviors occur (i.e., strategic VS. occasional behaviors). Trust ManagementService's Availability. A trust management service (TMS) provides an interface between users and cloud services for effective trust management. However, guaranteeing the availability of TMS is a difficult problem due to the unpredictable number of users and the highly dynamic nature of the cloud environment. Approaches that require understanding of users' interests and capabilities through similarity measurements or operational availability measurements (i.e., uptime to the total time) are inappropriate in cloud environments. TMS should be adaptive and highly scalable to be functional in cloud environments.

II. EXISTING AND PROPOSED SYSTEMS

A. Existing System

According to researchers at Berkeley, trust and security is ranked one of the top 10 obstacles for the adoption of cloud computing. Indeed, Service-Level Agreements (SLAs). Consumers' feedback is a good source to assess the overall trustworthiness of cloud services. Several researchers have recognized the significance of trust management and proposed solutions to assess and manage trust based on feedbacks collected from participants.

B. Proposed System

Cloud service users' feedback is a good source to assess the overall trustworthiness of cloud services. In this paper, we have presented novel techniques that help in detecting reputation based attacks and allowing users to effectively identify trustworthy cloud services. We introduce a credibility model that not only identifies misleading trust feedbacks from collusion attacks but also detects Sybil attacks no matter these attacks take place in a long or short period of time (i.e., strategic or occasional attacks respectively). We also develop an availability model that maintains the trust management service at a desired level. We also develop an availability model that maintains the trust management service at a desired level.

Advantages of Proposed System:

- Trust Cloud framework for accountability and trust in cloud computing. In particular, Trust Cloud consists of five layers including workflow,
- Propose a multi-faceted Trust Management (TM) system architecture for cloud computing to help the cloud service users to identify trustworthy cloud service providers.

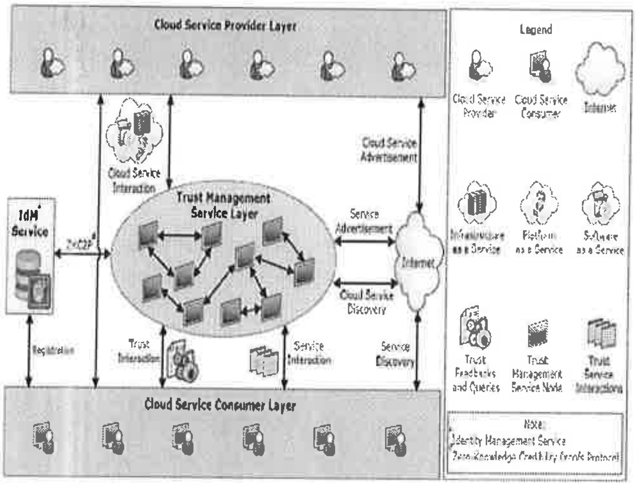

Cloud Service Provider Layer
Legend
Cloud Service Provider
Cloud Service Consumer
Internet
Cloud Service Advertisement
Cloud Service Interaction
Trust Management Service Layer
IdM* Service
ZKCCP#
Service Advertisement
Internet
Cloud Service Discovery
Infrastructure as a Service
Platform as a Service
Software as a Service
Registration
Trust Interaction
Service Interaction
Service Discovery
Trust Feedbacks and Queries
Trust Management Service Node
Trust Service Interactions
Cloud Service Consumer Layer
Note:
*Identity Management Service
#Zero-Knowledge Credibility Proofs Protocol

Fig.1. System Architecture.

III. METHODOLOGIES

A. Detection of Service

This layer consists of different users who use cloud services. For example, a new startup that has limited funding can consume cloud services. Interactions for this layer include: i) service discovery where users are able to discover new cloud services and other services through the Internet, ii) trust and service interactions where users are able to give their feedback or retrieve the trust results of a particular cloud service, and iii) registration where users establish their identity through registering their credentials in IdM before using TMS.

B. Trust Communication

In a typical interaction of the reputation based TMS, a user either gives feedback regarding the trustworthiness of a particular cloud service or requests the trust assessment of the service 1. From users' feedback, the trust behavior of a cloud service is actually a collection of invocation history records, represented by a tuple H=(C, S, F, T f), where C is the user's primary identity, S is the cloud service's identity, and F is a set of Quality of Service (QOS) feedbacks (i.e., the feedback represent several QOS parameters including availability, security, response time, accessibility, price).

C. IDM Registration

The system proposes to use the Identity Management Service (IdM) helping TMS in measuring the credibility of a consumer's feedback. However, processing the IdM information can breach the privacy of users. One way to preserve privacy is to use cryptographic encryption techniques. However, there is no efficient way to process encrypted data. Another way is to use anonymization techniques to process the IDM information without breaching the privacy of users. Clearly, there is a trade-off between high anonymity and utility.

D. Service Announcement and Communication

This layer consists of different cloud service providers who offer one or several cloud services, i.e., IaaS (Infrastructure as a Service), PaaS (Platform as a Service), and SaaS (Soft-ware as a Service), publicly on the Web (more details about cloud services models and designs can be found). These cloud services are accessible through Web portals and indexed on Web search engines such as Google, Yahoo, and Baidu. Interactions for this layer are considered as cloud service interaction with users and TMS.

IV. IMPLEMENTATION AND EXPERIMENTAL EVALUATION

In this section, we report the implementation and experimental results in validating the proposed approach. Our implementation and experiments were developed to validate and study the performance of both the credibility model and the availability model.

A. System Implementation

The trust management service's implementation is part of our large research project, named Cloud Armor, which offers a platform for reputation-based trust management of cloud services [10]. The platform provides an environment where users can give feedback and request trust assessment for a particular cloud service. Specifically, the trust management service (TMS) consists of two main components: the Trust Data Provisioning and the Trust Assessment Function.

1.The Trust Data Provisioning: This component is responsible for collecting cloud services and trust information. We developed the *Cloud Services Crawler* module based on the Open Source Web Crawler for Java (crawler4j) and extended it to allow the platform to automatically discover cloud services on the Internet. We implemented a set of functionalities to simplify the crawling process and made the crawled data more comprehensive (e.g., add Seeds (), select Crawling Domain (), add Crawling Time ()). In addition, we developed the Trust Feedbacks Collector module to collect feedbacks directly from users in the form of history records and stored them in the Trust Feedbacks Database: Indeed, users typically have to establish their identities for the first time they attempt to use the platform through registering their credentials at the Identity Management Service (IdM) which stores the credentials in the Trust

2. Identity Registry: Moreover, we developed the Identity Info Collector module to collect the total number of established identities among the whole identity behavior (i.e., all established identities for users who gave feedbacks to a particular cloud service).

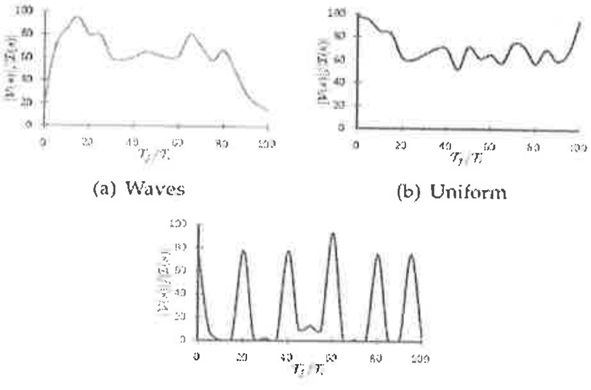
3. The Trust Assessment Function: This function is responsible for handling trust assessment requests from users where the trustworthiness of cloud services are compared and the factors of trust feedbacks are calculated (i.e., the credibility factors). We developed the Factors Calculator for attacks detection based on a set of factors (more details on how the credibility factors are calculated can be found). Moreover, we developed the Trust Assessor to compare the trustworthiness of cloud services through requesting the aggregated factors weights from the Factors Calculator to weigh feedbacks and then calculate the mean of all feedbacks given to each cloud service. The trust results for each cloud service and the factors' weights for trust feedbacks are stored in the Trust Results and Factors Weights Storage.

B. Experimental Evaluation

We particularly focused on validating and studying the robustness of the proposed credibility model against different malicious behaviors, namely collusion and Sybil attacks under several behaviors, as well as the performance of our availability model.

C. Credibility Model Experiments

We tested our credibility model using real world trust feedbacks on cloud services. In particular, we crawled several review websites such as cloud-computing.findthebest.com, cloudstorageprovidersreviews.com, and Cloud Hosting Reviewer.com, and where users give their feedbacks on cloud services that they have used. The collected data is represented in a tuple *H* where the feedback represents several QoS parameters as mentioned earlier and augmented with a set of credentials for each corresponding consumer. We managed to collect 10,076 feedbacks given by 6,982 users to 113 real-world cloud services. The collected dataset has been released to the research community via the project website. For experimental purposes, the collected data was divided into six groups of cloud services, three of which were used to validate the credibility model against collusion attacks, and the other three groups were used to validate the model against Sybil attacks where each group consists of 100 users. Each cloud service group was used to represent a different attacking behavior model, namely: *Waves*, *Uniform* and *Peaks* as shown in Fig.2. The behavior models represent the total number of malicious feedbacks introduced in a particular time instance (e.g., $|V(s)| = 60$ malicious feedbacks when $T_f = 40$, Fig.2(a)) when experimenting against collusion attacks. The behavior models also represent the total number of identities established by attackers in a period of time (e.g., $|I(s)| = 78$ malicious identities when $T_i = 20$, Fig.2(c)) where one malicious feedback is introduced per identity when experimenting against Sybil attacks. In collusion attacks, we simulated malicious feedback to increase trust results of cloud services (i.e., self-promoting attack) while in Sybil attacks we simulated malicious feedback to decrease trust results (i.e., slandering attack). To evaluate the robustness of our credibility model with respect to malicious behaviors (i.e., collusion and Sybil attacks), we used two experimental settings: I) measuring the robustness of the credibility model with a conventional model *Con(s, t_0, t)* (i.e., turning $C_r(c, s, t_0, t)$ to 1 for all trust feedbacks), and II) measuring the performance of our model using two measures namely *precision* (i.e., how well TMS did in detecting attacks) and *recall* (i.e., how many detected attacks are actual attacks). In our experiments, TMS started rewarding cloud services that had been affected by malicious behaviors when the attacks percentage reached 25% (i.e., $et(s) = 25\%$), so the rewarding process would occur only when there was a significant damage in the trust result. We conducted 12 experiments where six of which were conducted to evaluate the robustness of our credibility model against collusion attacks and the rest for Sybil attacks. Each experiment is denoted by a letter from A to F, as shown in Table 1.



(a) Waves

(b) Uniform

(c) Peaks

Fig.2. Attacking Behavior Models

TABLE I. Behavior Experimental Design

Malicious Behaviors	Experimental Setting	Waves	Uniform	Peaks
Collusion Attacks	*I*	*A*	*B*	*C*
	II	*A'*	*B'*	*C'*
Sybil Attacks	*I*	*D*	*E*	*F*
	II	*D'*	*E'*	*F'*

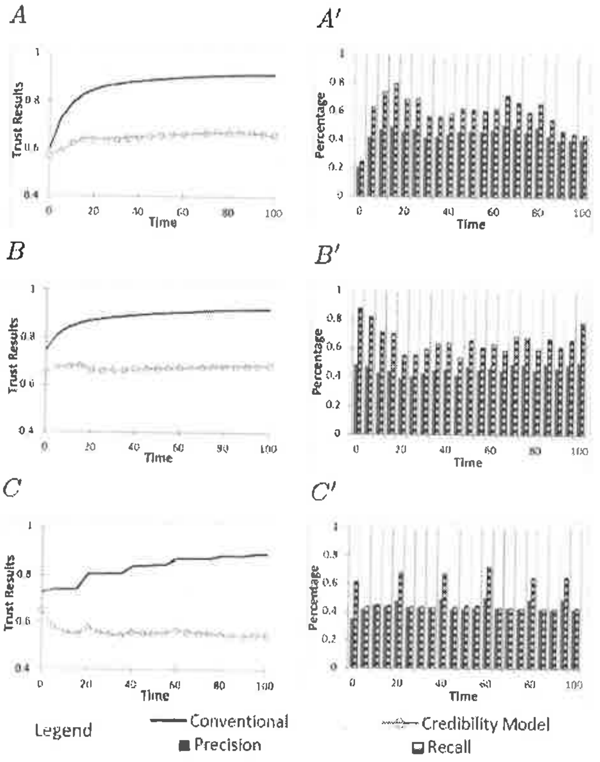


Fig.3. Robustness against Collusion Attacks.

1. Robustness against Collusion Attacks

For the collusion attacks, we simulated malicious users to increase trust results of cloud services (i.e., self promoting attack) by giving feedback with the range of [0.8, 1.0]. Fig.3 depicts the analysis of six experiments which were conducted to evaluate the robustness of our model with respect to collusion attacks. In Fig.3, *A*, *B*, and *C* show the trust result for experimental setting *I*, while *A'*, *B'*, and *C'* depict the results for experimental setting II. We note that the closer to 100 the time instance is, the higher the trust results are when the trust is calculated using the conventional model. This happens because malicious users are giving misleading feedback to increase the trust result for the cloud service. On the other hand, the trust results show nearly no change when calculated using the proposed credibility model (Fig.3 *A*, *B* and *C*). This demonstrates that our credibility model is sensitive to collusion attacks and is able to detect such malicious behaviors. In addition, we can make an interesting observation that our credibility model gives the best results in precision when the *Uniform* behavior model is used (i.e., 0.51, see Fig.3 *B'*), while the highest recall score is recorded when the *Waves* behavior model is used (i.e., merely 0.9, see Fig.3 *A'*). Overall, recall scores are fairly high when all behavior models are used which indicate that most of the detected attacks are actual attacks. This means that our model can successfully detect collusion attacks (i.e., whether the attack is strategic such as in *Waves* and *Uniform* behavior models or occasional such as in the *Peaks* behavior model) and TMS is able to dilute the increased trust results from self-promoting attacks using the proposed credibility factors.

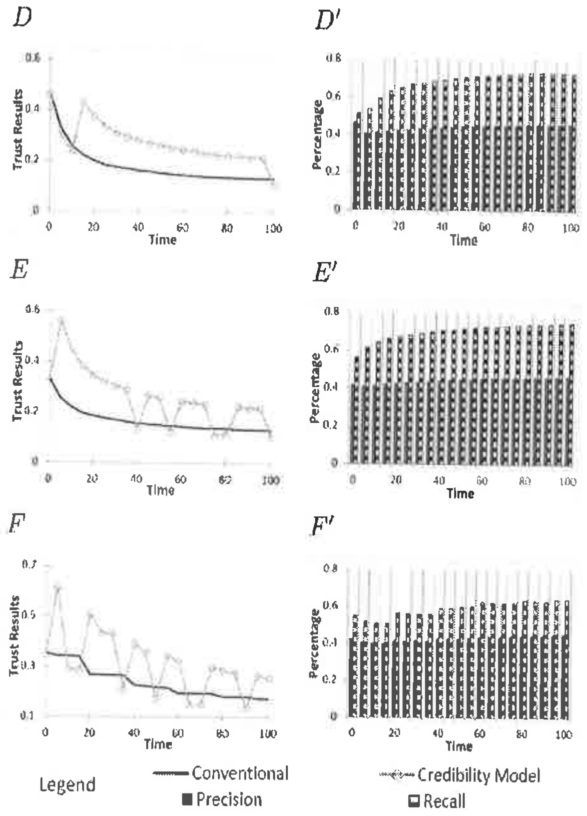
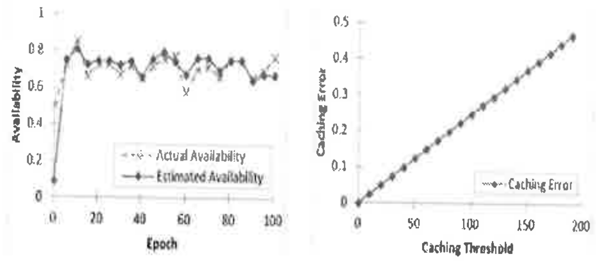


Fig.4. Robustness against Sybil Attacks.

2. Robustness against Sybil Attacks

For the Sybil attacks experiments, we simulated malicious users to decrease trust results of cloud services (i.e., slandering attack) by establishing multiple identities and giving one malicious feedback with the range of [0, 0.2] per identity. Fig.4 depicts the analysis of six experiments which were conducted to evaluate the robustness of our model with respect to Sybil attacks. In Fig.4, *D*, *E*, and *F* show the trust results for experimental setting *I*, while *D'*, *E'*, and *F'* depict the results for experimental setting II. From Fig.4, we can observe that trust results obtained by using the conventional model decrease when the time instance becomes closer to 100. This is because of malicious users who are giving misleading feedback to decrease the trust result for the cloud service. On the other hand, trust results obtained by using our proposed credibility model are higher than the ones obtained by using the conventional model (Fig.4 *D*, *E* and *F*). This is because the cloud service was rewarded when the attacks occurred. We also can see some sharp drops in trust results obtained by considering our credibility model where the highest number of drops is recorded when the *Peaks* behavior model is used (i.e., we can see 5 drops in Fig.4 *F* which actually matches the drops in the *Peaks* behavior model in Fig.2(c)). This happens because TMS will only reward the affected cloud services if the percentage of attacks during the same period of time has reached the threshold (i.e., which is set to 25% in this case). This means that TMS has rewarded the affected cloud service using the change rate of trust results factor.



(a) Actual Availability VS. Estimated Availability (b) Trust Results Caching Error Rate

Fig.5. Availability Prediction and Caching Accuracy

Moreover, from Fig.4 D', E' and F', we can see that our credibility model gives the best results in precision when the *Waves* behavior model is used (i.e., 0.47, see Fig.3 D'), while the highest recall score is recorded when the *Uniform* behavior model is used (i.e., 0.75, see Fig.3 A'). This indicates that our model can successfully detect Sybil attacks (i.e., either strategic attacks such as in *Waves* and *Uniform* behavior models or occasional attacks such as in the *Peaks* behavior model) and TMS is able to reward the affected cloud service using the change rate of trust results factor.

D. Availability Model Experiments

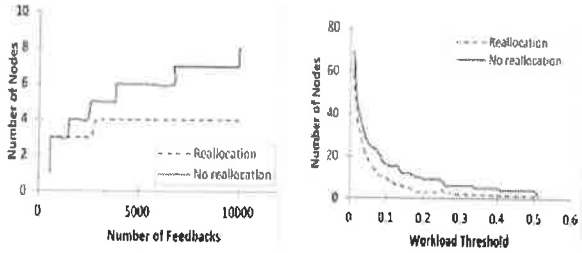
We tested our availability model using the same dataset we collected to validate the credibility model. However, for the availability experiments, we focused on validating the availability prediction accuracy, trust results caching accuracy, and reallocation performance of the availability model (i.e., to validate the three proposed algorithms including Particle Filtering based Algorithm, Trust Results & Credibility Weights Caching Algorithm, and Instances Management Algorithm).

1. Availability Prediction Accuracy

To measure the prediction accuracy of the availability model, we simulated 500 nodes hosting TMS instances and set the failure probability for the nodes as 3.5 percent, which complies with the findings. The motivation of this experiment is to study the estimation accuracy of our approach. We simulated TMS nodes' availability fluctuation and tracked their fluctuation of availability for 100 time steps (each time step counted as an *epoch*). The actual availability of TMS nodes and corresponding estimated availability using our particle filter approach were collected and compared. Fig.5 (a) shows the result of one particular TMS node. From the figure, we can see that the estimated availability is very close to the actual availability of the TMS node. This means that our approach works well in tracing and predicting the availability of TMS nodes.

2. Trust Results Caching Accuracy

To measure the caching accuracy of the availability model, we varied the caching threshold to identify the optimal number of new trust feedbacks that TMS received to recalculate the trust result for a particular cloud service without having a significant error in the trust results. The trust result caching accuracy is measured by estimating the root-mean-square error (RMSE) (denoted caching error) of the estimated trust result and the actual trust result of a particular cloud service. The lower the RMSE value means the higher accuracy in the trust result caching. Fig.5 (b) shows the trust result caching accuracy of one particular cloud service. From the figure, we can see that the caching error increases almost linearly when the caching threshold increases. The results allow us to choose the optimal caching threshold based on an acceptable caching error rate. For example, if 10% is an acceptable error margin, the caching threshold can be set to 50 feedbacks. It is worth mentioning that the caching error was measured on real users' feedbacks on real-world cloud services.



(a) Number of TMS Nodes VS. Feedbacks (b) Number of TMS Nodes VS. Workload Threshold

Fig.6. Reallocation Performance.

3. Reallocation Performance

To validate the reallocation performance of the availability model, we used two experimental settings: I) comparing the number of TMS nodes when using the reallocation of trust feedbacks and without reallocation while increasing the number of feedbacks (i.e., when the workload threshold e_w (s_{tms}) = 25%); II) comparing the number of TMS nodes when using the reallocation of trust feedbacks and without reallocation while varying e_w (s_{tms}). The lower the number of TMS nodes, the more cost efficient TMS is. Fig.6 (a) shows the results of experimental settings I. We can observe that the total number of TMS nodes when using the reallocation of trust feedbacks technique is fairly low and more stable than the total number of TMS nodes when reallocation is not used (i.e., even when the total number of feedbacks is high). Fig.6 (b) shows the results of experimental settings II. From the figure, we can see that the higher the workload threshold the lower the number of TMS nodes. However, the number of TMS nodes when using the reallocation of trust feedbacks technique is lower than the number of TMS nodes when reallocation is not considered. This means that our approach has advantages in minimizing the bandwidth cost by reducing the total number of TMS nodes.

V. CONCLUSION

From this Cloud Armor Supporting Reputation-based Trust Management for Cloud Services has been implemented. In cloud computing growth, the management of trust element is most challenging issue. Cloud computing has produce high challenges in security and privacy by the changing of environments. Trust is one of the most concerned obstacles for the adoption and growth of cloud computing. Although several solutions have been proposed recently in managing trust feedbacks in cloud environments, how to determine the credibility of trust feedbacks is mostly neglected. Additionally in future, we also enhance the performance of cloud as well as the security.

VI. REFERENCES

[1]Talal H. Noor, Quan Z. Sheng, Member, IEEE, Lina Yao, Schahram Dustdar, Senior Member, IEEE, and Anne H.H. Ngu, "CloudArmor: Supporting Reputation-based Trust Management for Cloud Services", IEEE Transactions on Parallel and Distributed Systems, Vol. 0, No. 0, 2014.

[2]S. M. Khan and K. W. Hamlen, "Hatman: Intra-Cloud Trust Management for Hadoop," in Proc. CLOUD'12, 2012.

[3]S. Pearson, "Privacy, Security and Trust in Cloud Computing," in Privacy and Security for Cloud Computing, ser. Computer Communications and Networks, 2013, pp. 3–42.

[4]J. Huang and D. M. Nicol, "Trust Mechanisms for Cloud Computing," Journal of Cloud Computing, vol. 2, no. 1, pp. 1–14, 2013.

[5]K. Hwang and D. Li, "Trusted Cloud Computing with Secure Resources and Data Coloring," IEEE Internet Computing, vol. 14, no. 5, pp. 14–22, 2010.

[6]M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A View of Cloud Computing," Communications of the ACM, vol. 53, no. 4, pp. 50–58, 2010.

[7]S. Habib, S. Ries, and M. Muhlhauser, "Towards a Trust Management System for Cloud Computing," in Proc. of TrustCom'11, 2011.

[8]I. Brandic, S. Dustdar, T. Anstett, D. Schumm, F. Leymann, and R. Konrad, "Compliant Cloud Computing (C3): Architecture and Language Support for User-Driven Compliance Management in Clouds," in Proc. of CLOUD'10, 2010.

[9]W. Conner, A. Iyengar, T. Mikalsen, I. Rouvellou, and K. Nahrstedt, "A Trust Management Framework for Service-Oriented Environments," in Proc. of WWW'09, 2009.

[10]T. H. Noor, Q. Z. Sheng, and A. Alfazi, "Reputation Attacks Detection for Effective Trust Assessment of Cloud Services," in Proc. of TrustCom'13, 2013.

[11]T. H. Noor, Q. Z. Sheng, S. Zeadally, and J. Yu, "Trust Management of Services in Cloud Environments: Obstacles and Solutions," ACM Computing Surveys, vol. 46, no. 1, pp. 12:1–12:30, 2013.

[12]S. Pearson and A. Benameur, "Privacy, Security and Trust Issues Arising From Cloud Computing," in Proc. Cloud Com'10, 2010.

Author's Profile:

Dasari Swapna, PG Scholar, Dept of CSE, Annamacharya Institute of Technology and Science, Hyderabad, TS, India, E-mail: dasariswapna912@gmail.com.

Mr.Ramesh Babu Varugu, received the Master of Technology degree in Information Technology from the Gurunanak Institute Of Science Andtechnology-JNTUH, he received the Bachelor of Technology degree from Lakireddy Balireddy College of Engineering, JNTUK. He is currently working as Associate Professor and a Head of the Department of CSE with Annamacharya Institute Of Technology And Sciences, Hyderabad. His interest subjects are operating Systems,Cloud Computing and etc. Email: ramesh.vnl@@gmail.com.

Mr.B.Ravinder Reddy, received the Master of Technology degree in Computer Science and Engineering from the Vasavya Engineering College-JNTUH, he received the Bachelor of Engineering degree from Jyothi Engineering College-JNTUH. He is currently working as Assistant Professor of CSE with Annamachraya Institute of Technology And Sciences, Hyderabad. His interest subjects are Data Mining, Software Engineering and etc.Email: ravibaireddy@gmil.com.

ijiTECH International Journal of Innovative Technologies
WWW.IJITECH.ORG
ISSN 2321-8665
Vol.04,Issue.13,
September-2016,
Pages:2304-2309

Energy-Aware Load Balancing and Application Scaling for the Cloud Ecosystem

A.SINDHUJA[1], RAMESH BABU VARUGU[2]
[1]PG Scholar, Dept of CSE. Annamacharya Institute of Technology and Sciences, Hyderabad, India,
E-mail: sindhuja.annam@gmail.com.
[2]HOD, Dept of CSE, Annamacharya Institute of Technology and Sciences, Hyderabad, India,
E-mail: ramesh.vnl@gmail.com.

Abstract: The energy consumption of computer and communication systems does not scale linearly with the workload. A system uses a significant amount of energy even when idle or lightly loaded. A widely reported solution to resource management in large data centers is to concentrate the load on a subset of servers and, whenever possible, switch the rest of the servers to one of the possible sleep states. We propose a reformulation of the traditional concept of load balancing aiming to optimize the energy consumption of a large-scale system: distribute the workload evenly to the smallest set of servers operating at an optimal energy level, while observing QoS constraints, such as the response time. Our model applies to clustered systems; the model also requires that the demand for system resources to increase at a bounded rate in each reallocation interval. In this paper we report the VM migration costs for application scaling.

Keywords: Load Balancing, Application Scaling, Idle Servers, Server Consolidation, Energy Proportional Systems.

I. INTRODUCTION

The concept of "load balancing" dates back to the time the first distributed computing systems were implemented in the late 1970s and early 1980s. It means exactly what the name implies, to evenly distribute the workload to a set of servers to maximize the throughput, minimize the response time, and increase the system resilience to faults by avoiding overloading one or more systems in the distributed environment. Distributed systems became popular after communication networks allowed multiple computing engines to effectively communicate with one another and the networking software became an integral component of an operating system. Once processes were able to easily communicate with one another using sockets1, the client-server paradigm became the preferred method to develop distributed applications; it enforces modularity, provides a complete isolation of clients from the servers, and enables the development of stateless servers. The client-server model proved to be not only enduring, but also increasingly successful; three decades later, it is at the heart of utility computing. In the last few years packaging computing cycles and storage and offering them as a metered service became a reality. Large farms of computing and storage platforms have been assembled and a fair number of Cloud Service Providers (CSPs) offer computing and storage services based on three different delivery models SaaS (Software as a Service), PaaS (Platform as a Service), and IaaS (Infrastructure as a Service). Reduction of energy consumption thus, of the carbon footprint of cloud related activities, is increasingly more important for the society. Indeed, as more and more applications run on clouds, more energy is required to support cloud computing than the energy required for many other human related activities. While most of the energy used by data centers is directly related to cloud computing, a significant fraction is also used by the networking infrastructure used to access the cloud. This fraction is increasing, as wireless access becomes more popular and wireless communication is energy intensive. In this paper we are only concerned with a single aspect of energy optimization, minimizing the energy used by cloud servers. Unfortunately, computer and communication systems are not energy proportional systems, in other words, their energy consumption does not scale linearly with the workload; an idle system consumes a rather significant fraction, often as much as 50%, of the energy used to deliver peak performance.

Cloud elasticity, one of the main attractions for cloud users, comes at a stiff price as the cloud resource management is based on over-provisioning. This means that a cloud service provider has to invest in a larger infrastructure than a "typical" or average cloud load warrants. At the same time, cloud elasticity implies that most of the time cloud servers operate with a low load, but still use a large fraction of the energy necessary to deliver peak performance. The low average cloud server utilization affects negatively the common measure of energy efficiency, the performance per Watt of power and amplifies the ecological impact of cloud computing. The strategy for resource management in a computing cloud we discuss is to concentrate the load on a subset of servers and, whenever possible, switch the rest of the servers to a sleep state. In a sleep state the energy consumption is very low. This observation implies that the traditional concept of load balancing could be reformulated to optimize the energy consumption of a large-scale system as follows: distribute evenly the workload to the smallest set of servers operating at an optimal energy level, while observing QoS constraints, such as the response time. An optimal energy level is one when the normalized system performance, defined as the ratio of the current performance to the maximum performance, is delivered with the minimum normalized energy consumption, defined as the ratio of the current energy consumption to the maximal one.

II. EXISTING AND PROPOSED SYSTEMS

A. Existing System

We also assume a clustered organization, typical for existing cloud infrastructure when the existing applications scale up above of the capacity with all servers running then the cluster leader interacts with the leaders of other clusters to satisfy the requests. This case is not addressed in this paper.

B. Proposed System

We introduce an energy-aware operation model used for load balancing and application scaling on a cloud. The basic philosophy of our approach is defining an energy-optimal operation regime and attempting to maximize the number of servers operating in this regime. Idle and lightly-loaded servers are switched to one of the sleep states to save energy. The load balancing and scaling algorithms also exploit some of the most desirable features of server consolidation mechanisms discussed in the literature.

III. MODULES

1. Load balancing,
2. Application scaling
3. Idle servers
4. Server consolidation,
5. Energy proportional systems.

A. Modules Description

1. Load Balancing

The concept of load balancing" dates back to the time when the first distributed computing systems were implemented. It means exactly what the name implies, to evenly distribute the workload to a set of servers to maximize the throughput, minimize the response time, and increase the system resilience to faults by avoiding overloading the systems.

2. Idle servers

Idle and under-utilized servers contribute significantly to wasted energy, see Section survey reports that idle servers contribute 11 million tons of unnecessary CO2 emissions each year and that the total yearly costs for idle servers is billion. An energy-proportional system consumes no energy when idle, very little energy under a light load, and gradually, more energy as the load increases.

3. Server Consolidation

The term server consolidation is used to describe: switching idle and lightly loaded systems to a sleep state; workload migration to prevent overloading of systems any optimization of cloud performance and energy efficiency by redistributing the workload discussed in Section For example, when deciding to migrate some of the VMs running on a server or to switch a server to a sleep state, we can adopt a conservative policy similar to the one advocated by auto scaling to save energy. Predictive policies, such as the ones discussed in will be used to allow a server to operate in a suboptimal regime when historical data regarding its workload indicates that it is likely to return to the optimal regime in the near future

4. Energy Proportional Systems

The energy efficiency of a system is captured by the ratio performance per Watt of power." During the last two decades the performance of computing systems has increased much faster than their energy efficiency

5. Energy Proportional Systems

In an ideal world, the energy consumed by an idle system should be near zero and grow linearly with the system load. In real life, even systems whose energy requirements scale linearly, when idle, use more than half the energy they use at full load. Data collected over a long period of time shows that the typical operating regime for data center servers is far from an optimal energy consumption regime. The dynamic range I s the deference between the upper and the lower limits of the energy consumption of a system as a function of the load placed on the system. A large dynamic range means that a system is able to operate at a lower fraction of its peak energy when its load is low

IV. SIMULATION EXPERIMENTS

The effectiveness of an energy-aware load balancing and scaling algorithm is characterized by its computational efficiency, the number of operations per Watt of power, and by its negative impact reacted by the potential SLA violations. In our study we shall use the data based on the measurements reported. These measurements are for a transaction processing benchmark, SPEC power ssj_2008, thus, the computational efficiency will be the number of transactions per Watt. We define the ideal computational efficiency as the number of transactions when all servers operate at the upper limit of the optimal region, at 80% load, and no SLA violations occur. In this case

$$\left(\frac{T}{P}\right)_{ideal} = \frac{T_{80\%}}{P_{80\%}} = \frac{1049}{235} = 4.46 \frac{transations}{Watt}. \qquad (1)$$

We conduct a simulation study to evaluate the effectiveness of the algorithms discussed. The simulation experiments reported in this paper were conducted on the Amazon cloud; a *c3.8xlarge* EC2 instance with 60G memory and 32 cores was used. The study will give us some indications about the operation of the algorithm in clusters of different sizes and subject to a range of workloads. The metrics for assessing the effectiveness and the overhead of the algorithms are:

- The evolution of the number of servers in each of the five operating regimes as a result of the load migration mandated by the algorithm.
- The computational efficiency before and after the application of the algorithm.
- The ratio of local versus in-cluster scaling decisions during simulation. This reacts the overhead of the algorithm.

For simplicity we chose only two sleep states C_3 and C_6 in the simulation. If the load of the cluster is more than 60% of the cluster capacity we do not choose the C_6 state because the probability that the system will require additional resources in the next future is high. Switching from the C_6 state to C_0 requires more energy and takes more time. On the other hand, when the cluster load is less than 60% of its capacity we choose the C_6 state because it is unlikely that the server will be reactivated in the next future. The simulation uses Amazon Web Services (AWS). AWS offers several classes of services; the servers in each class are characterized by the architecture, CPU execution rate, main memory, disk space, and I/O bandwidth the more powerful the server, the higher the cost per hour for the class of service. Table 1 summarizes the effects of application scaling and load balancing algorithm on a cluster when the parameters of the servers and the application is a transaction processing system. In a transaction processing system there is no workload migration, a front-end distributes the transactions to the servers thus, and we did not include migration costs. To assess the power consumption and the performance measured by the number of transactions we use average values for each regime. Recall that in this case the boundaries of the five operating regimes.

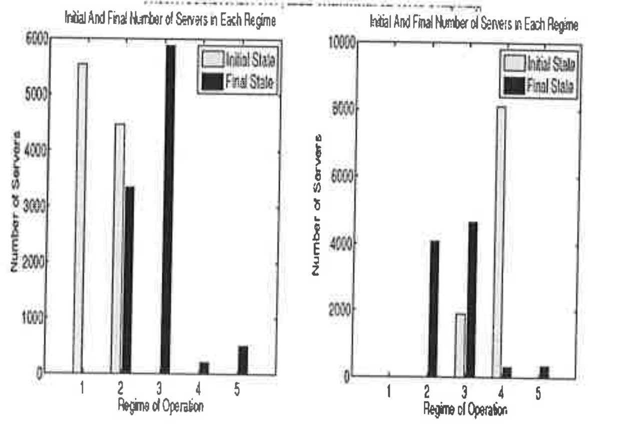
TABLE1. The Effects of Application Scaling and Load Balancing Algorithm on a System With the Parameters Described we Experimented with Two Average System Loads, 30% and 70% of the Server Capacity and Three Different Cluster Sizes, 10^2, 10^3, And 10^4. The Data Before/After the Application of the Algorithm Are: (a) The Number of Servers in Each One of the Five Operating Regimes, R_1,R_2,R_3,R_4,R_5 and in the Sleep State (Slp) (Columns 3-8); (b) P - The Average Power Consumption Per Processor in Watts(Column9);(c) T- The Performance Measured as the Average Number of Transactions Per Processor (Column 10); and (d) The Average Ratio T/P (Column 11).

Load	Size	# in R_1	# in R_2	# in R_3	# in R_4	# in R_5	# in slp	P	T	T/P
30%	10^2	65/0	35/27	0/64	0/2	0/3	0/4	188/197	264/746	1.4/3.8
	10^3	550/0	450/300	0/598	0/20	0/64	0/18	190/214	412/825	2.3/3.8
	10^4	5500/0	4500/3050	0/5950	0/50	0/319	0/631	190/203	442/771	2.3/3.8
70%	10^2	0/0	0/58	20/35	80/2	0/5	0/0	234/204	1,052/742	4.5/3.6
	10^3	0/0	0/430	190/490	810/20	0/56	0/4	231/215	1,054/823	4.5 /3.7
	10^4	0/0	0/4000	2000/4500	8000/250	0/233	0/17	235/193	1,052/715	4.5/3.7

The effect of the system load in we report on simulation experiments designed to evaluate the effectiveness of the algorithms for load balancing and energy optimization during application scaling. One of the questions we addressed was whether the system load has an effect on the resource management strategy to force the servers in a cluster to operate within the boundaries of the optimal regime. The experiments we report in this paper are for clusters with 10^2, 10^3, and 10^4 servers consisting of multiple racks of a WSC. For each cluster size we considered two load distributions:

- **Low Average Load**- an initial load uniformly distributed in the interval 20-40% of the server capacity the distribution of the number of servers in the five operating regimes for clusters with 10^4 servers, before and after load balancing; the distributions for 10^2 and 10^3 servers in a cluster are similar. When the average server load is 30% of their capacity, the algorithm is very effective; it substantially improves the energy efficiency to 3.8 from as low as 1.4. After load balancing, the number of servers in the optimal regime increases from 0 to about 60% and a fair number of servers are switched to the sleep state.
- **High Average Load** - initial server load uniformly distributed in the 60-80% of the server capacity. Fig.1 (b) Shows that when the average server load is 70% of their capacity, the average computational efficiency decreases from about 4.5 to 3.6 as many servers, about 80% of them, are forced from R_4 to the R_2 and R_3 regimes to reduce the possibility of SLA violations. Initially, no servers operated in the R_5 regime. In this experiment we did not use the anticipatory strategy and allowed servers to operate in the R_5 regime after load balancing; as a result, a small fraction of servers ended up in the R_5 regime. There is a balance between computational efficiency and SLA violations; the algorithm can be tuned to maximize computational efficiency or to minimize SLA violations according to the type of workload and the system management policies.

These results are consistent with the ones reported for smaller cluster sizes, 20, 40, 60, and 80 servers. This shows that the algorithms operate effectively for a wide range of cluster sizes and for lightly, as well as, heavily loaded systems.



(a) Cluster size: 10^4. Average load: 30%

(b) Cluster size 10^4. Average load: 70%

Fig.1. The Effect of Average Server Load on the Distribution of the Servers in the Five Operating Regimes, R_1, R_2, R_3, R_4 And R_5, Before and After Energy Optimization and Load Balancing. Average Load: (a) 30% and (b) 70%. Cluster Size: 10^4. Similar Results are Obtained for Cluster Size 10^2 and 10^3.

High-cost versus low-cost application scaling cloud elasticity allows an application to seamlessly scale up and down. In the next set of simulation experiments we investigate horizontal and vertical application scaling. Horizontal scaling requires the creation of additional VMs to run the application on lightly loaded servers. In-cluster, horizontal scaling incurs higher costs than vertical scaling for load balancing. The higher costs in terms of energy consumption and time are due to the communication with the leader to identify the potential targets and then to transport the VM image to one or more of them. Local, vertical scaling, allows the VM running an application to acquire additional resources from the local server; local vertical scaling has lower costs, but it is only feasible if the server has sufficient free capacity. The average ratio of high-cost in-cluster horizontal scaling to low-cost local vertical scaling for a transient period of 40 reallocation intervals are summarized in Table 2 and in Fig.2 (a), (b), (c), (d), and (e). When the average workload is 30% the algorithm works better; the larger the cluster, the more effectively the algorithm handles the application scaling as most decisions are done locally. After 40 reallocation intervals the ratios for the two average workload are 0.3 and 0.7, respectively, for cluster size 104; this shows that at higher workload VM migrations are more intense. The trends are different for the two cases, the ratio decreases in time when the average load is 30% and increases when the average load is 70%; we expect these trends to continue in the steady-state regime. We carried out measurements of computational efficiency on a system with a 3GHz Intel Core i7 with 8 cores and a solid state disk.

TABLE II. Average and Standard Deviation of In-Cluster to Local Decisions Ratio for 30% and 70% Average Server Load for Three Cluster Sizes, 10^2, 10^3, and 10^4.

Cluster size	Average load	Average ratio	Standard deviation	Average load	Average ratio	Standard deviation
10^2	30%	0.69	0.57	70%	0.51	0.89
10^3	30%	0.33	0.21	70%	0.58	0.92
10^4	30%	0.49	0.27	70%	0.53	0.98

The system was running under OS X Yosemite. The application used for our measurements runs under Xcode 6, an integrated development environment containing a suite of software tools for OS X and iOS. The application is I/O intensive, it reads a record from the secondary storage carries out some trivial computations and writes back the record. Xcode allowed us to measure the normalized performance and the corresponding normalized power consumption and thus, to calculate the computational efficiency at the boundaries of the five operating regimes described. Xcode reports only the power used by the cores running the application, so the average computational efficiency we were able to measure refers only to the processor, rather than the entire system. We then considered a cloud infrastructure consisting of 10, 000 servers identical with the system we have measured. We used the distribution of the servers in each of the five regimes in Table II to compute the computational efficiency before and after the application of the algorithm. The results are summarized in Table II. We see that the average computational efficiency decreases from 1.42 in R_2 to 1.245 in R_3, and 1.055 in R_4. As a result, the computational efficiency due to application of our algorithm increases only from 1.03 to 1.21 for the lightly loaded system and shows a modest increase from 1.09 to 1.18 for the heavy load case.

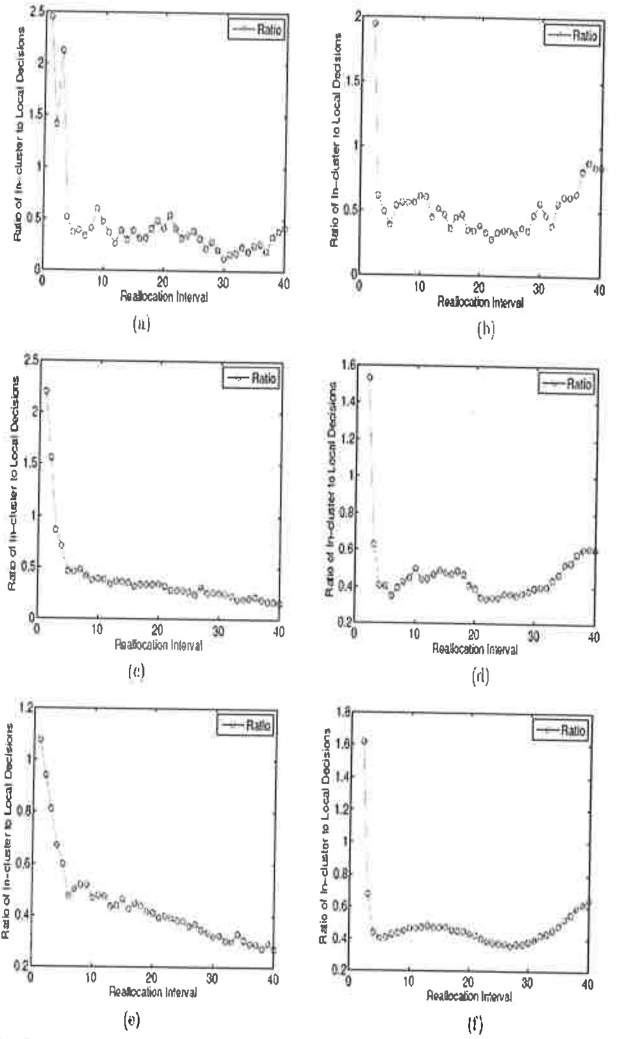


Fig.2. Time Series of In-Cluster to Local Decisions Ratios for a Transient Period of 40 Reallocation Intervals. Average Load: 30% of the Server Capacity in (a), (c), and (e); 70% in (b), (d), And (f). The Cluster Size: 10^2 In (a) and (b); 10^3 in (c) and (d); 10^4 in (e) and (f). After 40 Reallocation Intervals Almost Double Rations When the Load Is 70% Versus 30%.

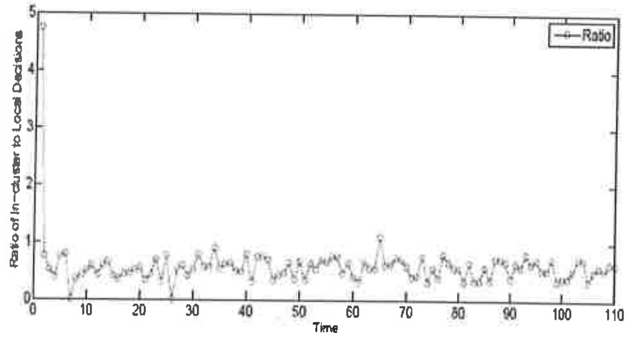


Fig.3. The Ratio of In-Cluster to Local Decisions in Response to Scaling Requests versus Time for a Cluster with 40 Servers When the Average Workload is 50% of the Server Capacity.

As noted in, the processors consume less than one third of their peak power at the very-low load thus, the actual improvement due to the application of our algorithm should be considerably higher. Finally, we attempted to carry out measurements in a realistic cloud environment and we investigated the possibility of using EC2 instances. First, we realized that no Mac OS based AMIs (Amazon Machine Images) are supported, so we could not use Xcode. We then discovered that the AWS virtual machine monitors prevent both Linux and Ubuntu AMIs to collect information related to power consumption. As a result, tools such as dmidecode or lshw used to report hardware information by monitoring kernel data structures return the fields related to power consumption as unknown. This lack of transparency makes the investigation of energy consumption in cloud environments a rather challenging task.

TABLE III. The Effects of Application Scaling and Load Balancing Algorithm on a System with 10^4 Servers 3ghz Intel Core I7. Shown are the Number of Servers in Each One of the Five Operating Regimes Before and After the Application of the Algorithm According to Table 1 and the Average Computational Efficiency in Each Regime $C_{ef}^{-R_i}$, $1 \leq I \leq 5$ Determined by Our Measurements. $\overline{C}_{ef}$Shows the Average Computational Efficiency Before and After the Application of the Algorithm.

Load	# in $\mathcal{R}_1$ $\bar{C}_{ef}^{\mathcal{R}_1} = 0.725$	# in $\mathcal{R}_2$ $\bar{C}_{ef}^{\mathcal{R}_2} = 1.420$	# in $\mathcal{R}_3$ $\bar{C}_{ef}^{\mathcal{R}_3} = 1.245$	# in $\mathcal{R}_4$ $\bar{C}_{ef}^{\mathcal{R}_4} = 1.055$	# in $\mathcal{R}_5$ $\bar{C}_{ef}^{\mathcal{R}_5} = 1.050$	C_{ef}
30%	5500/0	4500/3050	0/5950	0/50	0/319	1.03/1.21
70%	0/0	0/4000	2000/4500	8000/250	0/233	1.09/1.18

V. CONCLUSION

The average server utilization in large data-centers is 18%. When idle the servers of a data center use more than half the power they use at full load. The alternative to the wasteful resource management policy when the servers are always on, regardless of their load, is to develop energy-aware load balancing policies. Such policies combine dynamic power management with load balancing. There are ample opportunities to reduce the energy necessary to power the servers of a large-scale data center and shrink the carbon footprint of cloud computing activities, even though this is only a fraction of the total energy required by the ever increasing appetite for computing and storage services. To optimize the resource management of large farms of servers we redefine the concept of load balancing and exploit the technological advances and the power management functions of individual servers. In the process of balancing the load we concentrate it on a subset of servers and, whenever possible, switch the rest of the servers to a sleep state. From the large number of questions posed by energy-aware load balancing policies we discuss only the energy costs for migrating a VM when we decide to either switch a server to a sleep state or force it to operate within the boundaries of an energy optimal regime. The policies analyzed in this paper aim to keep the servers of a cluster within the boundaries of the optimal operating regime.

After migrating the VMs to other servers identified by the cluster leader, a lightly loaded server is switched to one of the sleep states. There are multiple sleep states; the higher the state number, the larger the energy saved, and the longer the time for the CPU to return to the state C_0 which corresponds to a fully operational system. For simplicity we chose only two sleep states C_3 and C_6 in the simulation. If the overall load of the cluster is more than 60% of the cluster capacity we do not switch any server to a C_6 state because in the next future the probability that the system will require additional computing cycles is high. Switching from the C_6 state to C_0 requires more energy and takes more time. On the other hand, when the total cluster load is less than 60% of its capacity we switch to C_6 because it is so unlikely that for the next interval and the interval after that system needs extra computational unit. The simulation results reported that the load balancing algorithms are effective and that low-cost vertical scaling occurs even when a cluster operates under a heavy load. The larger the cluster size the lower the ratio of high cost in-cluster versus low-cost local decisions. The QoS requirements for the three cloud delivery models are different thus; the mechanisms to implement a cloud resource management policy based on this idea should be different. To guarantee real-time performance or a short response time, the servers supporting SaaS applications such as data streaming or on-line transaction processing (OLTEP) may be required to operate within the boundaries of a sub-optimal region in terms of energy consumption. There are cases when the instantaneous demand for resources cannot be accurately predicted and systems are forced to operate in a non-optimal region before additional systems can be switched from a sleep state to an active one. Typically, PaaS applications run for extended periods of time and the smallest set of serves operating at an optimal power level to guarantee the required turnaround time can be determined accurately. This is also true for many IaaS applications in the area of computational science and engineering. There is always a price to pay for an additional functionality of a system, so the future work should evaluate the overhead and the limitations of the algorithms required by these mechanisms.

VI. REFERENCES

[1]Ashkan Paya and Dan C. Marinescu, "Energy-aware Load Balancing and Application Scaling for the Cloud Ecosystem", IEEE Transactions on Cloud Computing.

[2]D. Ardagna, B. Panicucci, M. Trubian, and L. Zhang. \Energy-aware autonomic resource allocation in multitier virtualized environments." IEEE Trans. on Services Computing, 5(1):2-19, 2012.

[3]J. Baliga, R.W.A. Ayre, K. Hinton, and R.S. Tucker. \Green cloud computing: balancing energy in processing, storage, and transport." Proc. IEEE, 99(1):149-167, 2011.

[4]L. A. Barroso and U. H□bzle. \The case for energy proportional computing." IEEE Computer, 40(12):33-37, 2007.

[5]L. A. Barossso, J. Clidaras, and U.H□bzle. The Data-center as a Computer; an Introduction to the Design of

Warehouse-Scale Machines. (Second Edition). Morgan & Claypool, 2013.

[6]A. Beloglazov, R. Buyya \Energy efficient resource management in virtualized cloud data centers." Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Comp., 2010.

[7]A. Beloglazov, J. Abawajy, R. Buyya. \Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing." Future Generation Computer Systems, 28(5):755-768, 2012.

[8]A. Beloglazov and R. Buyya. \Managing overloaded hosts for dynamic consolidation on virtual machines in cloud centers under quality of service constraints." IEEE Trans. on Parallel and Distributed Systems, 24(7):1366-1379, 2013.

[9]M. Blackburn and A. Hawkins. \Unused server survey resultsanalysis."www.thegreengrid.org/media/WhitePapers/Unused% 20Server%20Study WP 101910 v1. ashx?lang=en (Accessed on December 6, 2013).

[10]M. Elhawary and Z. J. Haas. \Energy-efficient protocol for cooperative networks." IEEE/ACM Trans. on Networking, 19(2):561-574, 2011.

[11]A. Gandhi, M. Harchol-Balter, R. Raghunathan, and M.Kozuch. \AutoScale: dynamic, robust capacity management for multi-tier data centers." ACM Trans. On Computer Systems, 30(4):1-26, 2012.

[12]A. Gandhi, M. Harchol-Balter, R. Raghunathan, and M.Kozuch. \Are sleep states effective in data centers?" Proc. Int. Conf. on Green Comp., pp. 1-10, 2012.

Author's Profile:

A. Sindhuja, PG Scholar, Department of CSE, From Annamacharya Institute of Technology and Sciences, Hyderabad, India, E-mail: sindhuja.annam@gmail.com.

Mr.Ramesh Babu Varugu received the Master of Technology degree in Information Technology from the Gurunanak Institute of Science And Technology-JNTUH, he received the Bachelor Of Engineering degree from Lakireddy Balireddy College Of Enngineering-JNTU-K. He is currently working as Associate Professor and a Head of the Department of CSE with Annamacharya institute of technology and sciences. His interest subjects are Operating Systems, Cloud computing.E-mail: ramesh.vnl@gmail.com.

IJSETR
International Journal of
Scientific Engineering and Technology Research
WWW.IJSETR.COM

RAMESH CHOWDARY. G
MANAGING EDITOR

21-01-2015

PAPER ID: SG-IJSETRV04IS02P4007

B. SATISH GUPTA,
PG Scholar, Dept of CSE,
Vizag Institute of Technology,
Visakhapatnam, AP, India.

Dear Author,

This is to inform you regarding the selection of your paper entitled **"Cloud Data Center Management with Quality of Service using a Novel Stochastic Mode"** after peer review in refereed International Journal of Scientific Engineering and Technology Research, **ISSN 2319-8885, Volume No.04, Issue No.02, January-2015.** This issue published by the **IJSETR.**

The hard copy of the journal shall be delivered to you soon after its release.

Please communicate to us for any further queries.

Regards,

Ramesh Chowdary.G
Managing Editor
IJSETR,
Hyderabad, INDIA.
Mobile: +91-9290860984
Email:ijournals@ijsetr.com.
Website: www.ijsetr.com.

Cloud Data Center Management With Quality Of Service Using A Novel Stochastic Mode

[1]B.SATISH GUPTA, [2]SHAFIULILAH SHAIK
[1]M.Tech Scholar, Dept. of CSE, Vizag Institute Of Technology,Visakhapatnam,AP, India
[2]Assistant Professor, Dept. of CSE, Vizag Institute Of Technology,Visakhapatnam,AP, India

Abstract-- Cloud data center management is a key problem due to the numerous and heterogeneous strategies that can be applied, ranging from the VM placement to the federation with other clouds. Performance evaluation of Cloud Computing infrastructures is required to predict and quantify the cost-benefit of a strategy portfolio and the corresponding Quality of Service (QoS) experienced by users. Such analyses are not feasible by simulation or on-the-field experimentation, due to the great number of parameters that have to be investigated. In this paper, we present an analytical model, based on Stochastic Reward Nets (SRNs), that is both scalable to model systems composed of thousands of resources and flexible to represent different policies and cloud-specific strategies. Several performance metrics are defined and evaluated to analyze the behavior of a Cloud data center: utilization, availability, waiting time, and responsiveness. A resiliency analysis is also provided to take into account load bursts. Finally, a general approach is presented that, starting from the concept of system capacity, can help system managers to opportunely set the data center parameters under different working conditions.

Index Terms-- Smart antenna; wireless mobile network; adaptive antenna; broadcast scheduling.

I.INTRODUCTION

In order to integrate business requirements and application level needs, in terms of Quality of Service (QoS), cloud service provisioning is regulated by Service Level Agreements (SLAs): contracts between clients and providers that express the price for a service, the QoS levels required during the service provisioning, and the penalties associated with the SLA violations. In such a context, performance evaluation plays a key role allowing system managers to evaluate the effects of different resource management strategies on the data center functioning and to predict the corresponding costs/benefits.

Cloud systems differ from traditional distributed systems. First of all, they are characterized by a very large number of resources that can span different administrative domains. Moreover, the high level of resource abstraction allows to implement particular resource management techniques such as VM multiplexing or VM live migration that, even if transparent to final users, have to be considered in the design of performance models in order to accurately understand the system behavior. Finally, different clouds, belonging to the same or to different organizations, can dynamically join each other to achieve a common goal, usually represented by the optimization of resources utilization. This mechanism, referred to as cloud *federation*, allows to provide and release resources on demand thus providing elastic capabilities to the whole infrastructure. On-the-field experiments are mainly focused on the offered QoS, they are based on a black box approach that makes difficult to correlate obtained data to the internal resource management strategies implemented by the system provider. Also it is interesting to note that Simulation does not allow to conduct comprehensive analyses of the system performance due to the great number of parameters that have to be investigated. In this paper, we present a stochastic model, based on Stochastic Reward Nets (SRNs), that exhibits the above mentioned features allowing to capture the key concepts of an IaaS cloud system. The proposed model is scalable enough to represent systems composed of thousands of resources and it makes possible to represent both physical and virtual resources exploiting cloud specific concepts such as the infrastructure elasticity. With respect to the existing literature, the innovative aspect of the present work is that a generic and comprehensive view of a cloud system is presented. Low level details, such as VM multiplexing, are easily integrated with cloud based actions such as federation, allowing to investigate different mixed strategies. An exhaustive set of performance metrics are defined regarding both the system provider (e.g., utilization) and the final users (e.g., responsiveness). Considering various advantages like to provide a fair comparison among different resource management strategies, also taking into account the system elasticity, a performance evaluation approach is described. Such an approach, based on the concept of system capacity, presents a holistic view of a cloud system and it allows system managers to study the better solution with respect to an established goal and to opportunely set the system parameters.

H. Liu [1] proposed that live migration of virtual machines (VM) across physical hosts provides a significant new benefit for administrators of data centers and clusters. Previous memory-to-memory approaches demonstrate the effectiveness of live VM migration in local area networks (LAN), but they would cause a long period of downtime in a wide area network (WAN) environment. This paper describes the design and implementation of a novel approach, namely, CR/TR-Motion, which adopts checkpointing/recovery and trace/replay technologies to provide fast, transparent VM migration for both LAN and WAN environments. With execution trace logged on the source host, a synchronization algorithm is performed to orchestrate the running source and target VMs until they reach a consistent state. CR/TR-Motion can greatly reduce the migration downtime and network bandwidth consumption. Experimental results show that the approach can drastically reduce migration overheads compared with memory-to-memory approach in a LAN: up to 72.4 percent on application observed downtime, up to 31.5 percent on total migration time, and up to 95.9 percent on the data to synchronize the VM state. The application performance overhead due to migration is kept within 8.54 percent on average. The results also show that for a variety of workloads migrated across WANs, the migration downtime is less than 300 milliseconds.

R. Buyya et.al [2] discussed that cloud computing aims to power the next generation data centers and enables application service providers to lease data center capabilities for deploying applications depending on user QoS (Quality of Service) requirements. Cloud applications have different composition, configuration, and deployment requirements. Quantifying the performance of resource allocation policies and application scheduling algorithms at finer details in Cloud computing environments for different application and service models under varying load, energy performance (power consumption, heat dissipation), and system size is a challenging problem to tackle. To simplify this process, in this paper we propose CloudSim: an extensible simulation toolkit that enables modelling and simulation of Cloud computing environments. The CloudSim toolkit supports modelling and creation of one or more virtual machines (VMs) on a simulated node of a Data Center, jobs, and their mapping to suitable VMs. It also allows simulation of multiple Data Centers to enable a study on federation and associated policies for migration of VMs for reliability and automatic scaling of applications.

A. Iosupet.al [3] propsoed that cloud computing is an emerging infrastructure paradigm that promises to eliminate the need for companies to maintain expensive computing hardware. Through the use of virtualization and resource time-sharing, clouds address with a single set of physical resources a large user base with diverse needs. Thus, clouds have the potential to provide their owners the benefits of an economy of scale and, at the same time, become an alternative for both the industry and the scientific community to self-owned clusters, grids, and parallel production environments. For this potential to become reality, the first generation of commercial clouds need to be proven to be dependable. In this work we analyze the dependability of cloud services. Towards this end, we analyze long-term performance traces from Amazon Web Services and Google App Engine, currently two of the largest commercial clouds in production. We find that the performance of about half of the cloud services we investigate exhibits yearly and daily patterns, but also that most services have periods of especially stable performance. Last, through trace-based simulation we assess the impact of the variability observed for the studied cloud services on three large-scale applications, job execution in scientific computing, virtual goods trading in social networks, and state management in social gaming. We show that the impact of performance variability depends on the application, and give evidence that performance variability can be an important factor in cloud provider selection.

II. PROPOSED SYSTEM ARCHITECTURE

The description of the proposed model is given as shown in the fig.1. The architecture description is mentioned with various blocks like various arrival processes, system queuing of the arrival jobs, scheduling and logical resources.

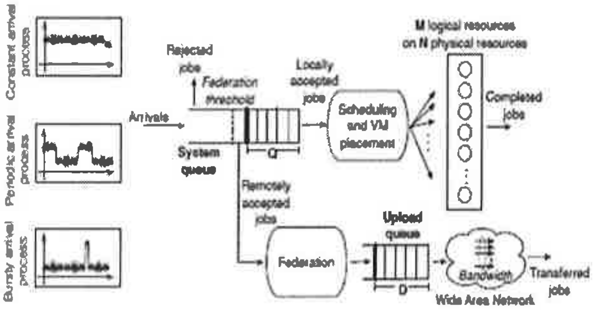


Fig.1 System Architecture

MODULES DESCRIPTION

a. System Queuing:

Job requests (in terms of VM instantiation requests) are en-queued in the system queue. Such a queue has a finite size Q, once its limit is reached further requests are rejected. The system queue is managed according to a FIFO scheduling policy.

b. Scheduling Module:

When a resource is available a job is accepted and the corresponding VM is instantiated. We assume that the instantiation time is negligible and that the service time (i.e.,

IV. PROCESS DIAGRAMS

USE CASE DIAGRAM

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

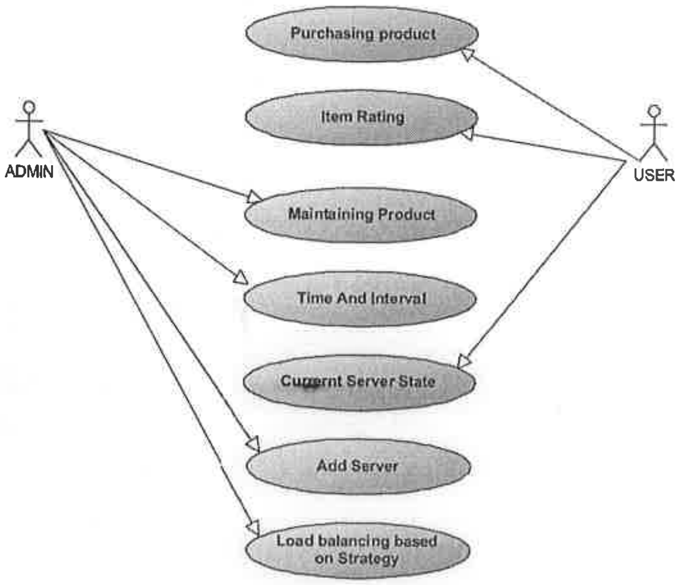


Fig.3. Use case diagram

CLASS DIAGRAM

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

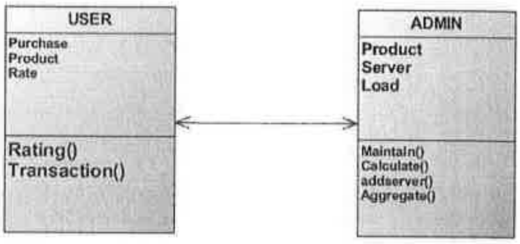


Fig.4. Class diagram

SEQUENCE DIAGRAM

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

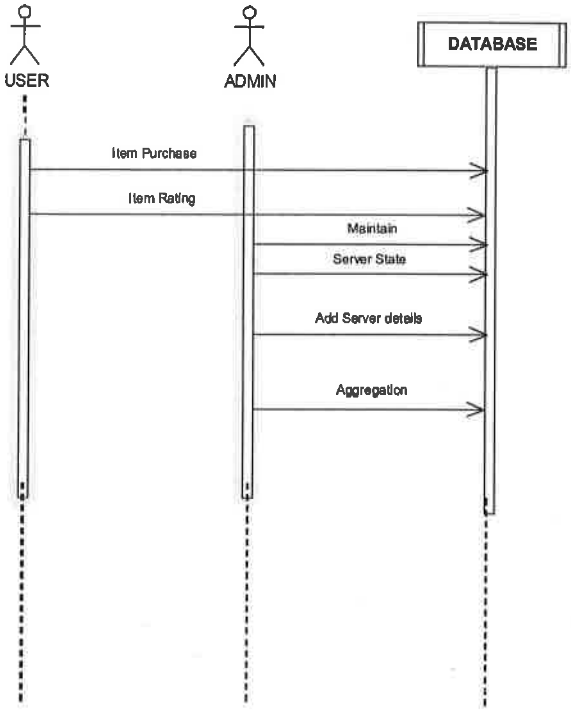


Fig.5. Sequence diagram

ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

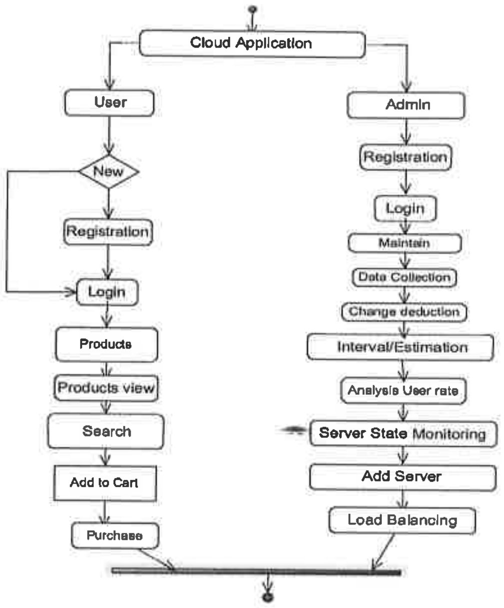


Fig.6. Activity diagram

V. RESULTS

Results pertaining to the proposed model are mentioned here as screen shots of the implementation.

STOCHASTIC MODEL TO INVESTIGATE DATA CENTER PERFORMANCE & QOS IN IAAS CLOUD

(a) Index page

STOCHASTIC MODEL TO INVESTIGATE DATA CENTER PERFORMANCE & QOS IN IAAS CLOUD

(b) Admin page

STOCHASTIC MODEL TO INVESTIGATE DATA CENTER PERFORMANCE & QOS IN IAAS CLOUD

(c) Account details

STOCHASTIC MODEL TO INVESTIGATE DATA CENTER PERFORMANCE & QOS IN IAAS CLOUD

(d) Server details

STOCHASTIC MODEL TO INVESTIGATE DATA CENTER PERFORMANCE & QOS IN IAAS CLOUD

(e) Server details editing

VI. CONCLUSION

In this paper, we have presented a stochastic model to evaluate the performance of an IaaS cloud system. Several performance metrics have been defined, such as availability, utilization, and responsiveness, allowing to investigate the impact of different strategies on both provider and user point-of-views. In a market-oriented area, such as the Cloud Computing, an accurate evaluation of these parameters is required in order to quantify the offered QoS and opportunely manage SLAs. Future works will include the analysis of autonomic techniques able to change on-thefly the system configuration in order to react to a change on the working conditions. We will also extend the model in order to represent PaaS and SaaS Cloud systems and to integrate the mechanisms needed to capture VM migration and data center consolidation aspects that cover a crucial role in energy saving policies.

REFERENCES

[1] R. Buyya *et al.*, "Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Gener. Comput. Syst.*, vol. 25, pp. 599–616, June 2009.

[2] X. Meng *et al.*, "Efficient resource provisioning in compute clouds via vm multiplexing," in *Proceedings of the 7th international conference on Autonomic computing*, ser. ICAC '10. New York, NY, USA: ACM, 2010, pp. 11–20.

[3] H. Liu *et al.*, "Live virtual machine migration via asynchronous replication and state synchronization," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 22, no. 12, pp. 1986 –1999, dec. 2011.

[4] B. Rochwerger *et al.*, "Reservoir - when one cloud is not enough," *Computer*, vol. 44, no. 3, pp. 44 –51, march 2011.

[5] R. Buyya, R. Ranjan, and R. Calheiros, "Modeling and simulation of scalable cloud computing environments and the cloudsim toolkit: Challenges and opportunities," in *High Performance Computing Simulation, 2009. HPCS '09. International Conference on*, june 2009, pp. 1 –11.

[6] A. Iosup, N. Yigitbasi, and D. Epema, "On the performance variability of production cloud services," in *Cluster, Cloud and Grid Computing (CCGrid), 2011 11th IEEE/ACM International Symposium on*, may 2011, pp. 104 –113.

[7] V. Stantchev, "Performance evaluation of cloud computing offerings," in *Advanced Engineering Computing and Applications in Sciences, 2009. ADVCOMP '09. Third International Conference on*, oct. 2009, pp. 187 – 192.

[8] S. Ostermann *et al.*, "A Performance Analysis of EC2 Cloud Computing Services for Scientific Computing," in *Cloud Computing*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer Berlin Heidelberg, 2010, vol. 34, ch. 9, pp. 115–131.

[9] H. Khazaei, J. Misic, and V. Misic, "Performance analysis of cloud computing centers using m/g/m/m+r queuing systems," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 23, no. 5, pp. 936 –943, may 2012.

[10] R. Ghosh, K. Trivedi, V. Naik, and D. S. Kim, "End-to-end performability analysis for infrastructure-as-a-service cloud: An interacting stochastic models approach," in *Dependable Computing (PRDC), 2010 IEEE 16th Pacific Rim International Symposium on*, dec. 2010, pp. 125 – 132.

[11] G. Ciardo *et al.*, "Automated generation and analysis of Markov reward models using stochastic reward nets." *IMA Volumes in Mathematics and its Applications: Linear Algebra, Markov Chains, and Queueing Models*, vol. 48, pp. 145–191, 1993.

[12] D. Gupta, L. Cherkasova, R. Gardner, and A. Vahdat, "Enforcing performance isolation across virtual machines in xen," in *Proceedings of the ACM/IFIP/USENIX International Conference on Middleware*, New York, NY, USA: Springer-Verlag New York, Inc., 2006, pp. 342–362.

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 2, Issue 2, February 2014

ORDERING AND SUGGESTING POPULAR ITEMSETS IN PHARMACY USING MODIFIED APRIORI ALGORITHM

A.Nagasri[1], B.Sujatha[2], CH.Bhavani[3], B.Ravinder reddy[4], B.Sathish Guptha[5]

Abstract: We considered the problem of ranking the popularity of items and suggesting popular items based on user feedback. User feedback is obtained by iteratively presenting a set of suggested items, and users selecting items based on their own preferences either the true popularity ranking of items, and suggest true popular items. We consider Apriori algorithm with some modifications overcoming the complexity that has been seen in other randomized algorithms. The most effective feature of this approach is that it reduces the number of database scans and complexity.

I.INTRODUCTION

1.1. TERMINOLOGY

In this section we first want to introduce the different terms that we were going to use in our paper as fallows.

1.1.1 Ranking: Ranking is giving rank scores to the most popular item by taking user feedback. The most frequently occurring item is given the highest rank score.

1.1.2 Selection: We focus on the ranking of items where the only available information is the observed selection of items. In learning of the users preference over items, one may leverage some side information about items, but this is out of the scope of this paper.

1.1.3 Imitate: The user study was conducted in very famous pharmacies and which has been used to set of tablets. The user may check the list and select the set of tablets which they like most and depending on those like results the new suggestion list has been developed by the algorithm.

1.1.4 Popular: In practice, one may use prior information about item popularity. For example, in the survey the user may select the suggested tablet or they may also select the others. If they selected the already suggested items they will become more popular and if he doesn't they may get out of the popular list.

1.1.5 Association Rule: Association Rules are if/then statements that help uncover relationships between seemingly unrelated data in the relational database or other information repository. An example of an association rule would be if a customer buys a cold tablet, he is 60% interested in also purchasing cough tablets.

II. THEORETICAL STUDY

We consider the cold tablet selection and suggesting the best sold cold tablet and their combinations that were most liked by most of the users. Consider a set of cold tablets M: (m1, m2, m3, m4, ….mn) where n > 1. Now we were calculating the set of items in C where were mostly sold and mostly liked by the users, as S
S: (s1, s2, s3, s4, …. sg) where g > 1.

We need to consider an item I, we interpret si as the portion of users that would select item i if suggestions were not made. We assume that the popularity rank scores s as follows:
a) Items of set S were estimated to is as $s1 \geq s2 \geq s3 \geq \ldots sc$,
b) s is completely normalized such that it is a probabilitydistribution, i.e., s1 + s2 + s3 + …. +sc = 1. c) si is always positive for all items i.

III. PROPOSED ALGORITHM AND STUDY

We have some of the systems already existing in the same field and we have also identified some of the disadvantages in them as follows:

- The popularity for any item is given based on the production of that item. This may not give good result because customers may not have the knowledge of true popularity they needed and depend on the results given by the producer.
- The updates are performed regardless of the true popularity by virtual analysis.
- Producer have to analyses things manually and complexity involves in this. Due to this time consumption may be high.
- The algorithms used in this system may fail to achieve true popularity.

We consider the problem learning of the popularity of items that is assumed to be unknown but has to be learned from the observed user's selection of items. We have selected a mobile market and mobile distribution outlets as our data set and examined them completely in all areas where we can give the list of items suggested by the users and we have made web-application to make an survey at real-time and considered the data given by more that 1000 members of different categories of people and applied our proposed modified apriori algorithm on the set of data and started suggesting the item in the mobile outlets for the actual users, which had helped the mobile phone companies and also the outlet in-charges. We have implemented the same in some of the mobile outlets in INDIA where we got very good results. The actual goal of the system is to efficiently learn the popularity of items and suggest the popular items to users. This was done to the user to suggest them the mostly used mobiles and their accessories, such that they also buy the best and at the same time the outlet owner will also get benefited. The most important feature in our project is suggesting the users by refreshing the latest results every time the user gives the input and changes his like list.

Now we have overcome many of the disadvantages of the existing systems and achieved many advantages with the proposed algorithm and method as follows:

- In our approach, we consider the problem of ranking the popularity of items and suggesting popular items based on user feedback.
- User feedback is obtained by iteratively presenting a set of suggested items, and users selecting items based on their own preferences either from this suggestion set or from the set of all possible items.
- The goal is to quickly learn the true popularity ranking of items and suggest true popular items.
- In this system better algorithms are used. The algorithms use ranking rules and suggestion rules in order to achieve true popularity.

IV. PROPOSED ALGORITHM APRIORI ALGORITHM

This is to find frequent item-sets using candidate generation. Apriori employs an iterative approach known as level-wise search, where k-itemsets are used to explore (k+1) itemsets. First, the set of frequent 1-itemsets is found by scanning database to accumulate the count for each item and collecting those items that satisfy minimum support. The resulting set is denoted by L1. Next, L1 is used to find L2,the set of frequent 2-itemsets,which is use to find L3 and so on, until no more frequent k-itemsets can be found. The finding of each Lk requires one full scan of database. To improve the efficiency of the level-wise generation of frequent itemsets, an important property is called apriori property, which is used to reduce search space.

Apriori property:

- All non empty subsets of a frequent itemset must also be frequent.
- Antimonotone property-if a set cannot pass a test, all of its supersets will fail the same test as well.

Apriori algorithm is a two step process, the two steps are

- Join step
- Prune step

Join step:
To find Lk, a set of candidate k-itemsets is generated by joining Lk-1 with itself. This set of candidates is denoted by Ck. Let l1 and l2 be itemsets in Lk-1.The notation li[j] refers to jth item in li. By convention, Apriori assumes that items within a transaction or itemset are sorted in lexicographic order. For the (k-1) itemset, li, this means that the items are sorted such that li[1]<l1[2]<..........<li[k-1].The join Lk-1∞Lk-1, is performed, where members of Lk-1 are joinable if their first (k-2) items are in common.

Prune step:
Ck is a superset of Lk, that is, its members may or may not be frequent but all of frequent k-itemsets are included in Ck. A scan of database to determine the count of each candidate in Ck would result in the determination of Lk. Ck however can be huge. so this could involve heavy computation. To reduce the size of Ck, Apriori property is used as follows: Any (k-1) itemset that is not frequent cannot be a subset of a frequent k-itemset. Hence, if any (k-1) subset of a candidate k-itemset is not in Lk-1, then the candidate cannot be frequent either and so can be removed from Ck. This subset testing can be quickly done by maintaining a hash tree of all frequent itemsets.

Improving the efficiency of apriori by proposing it with some variations. Several variations are summarized as follows:

- Hash-based technique (hashing itemsets into corresponding buckets):A hash-based technique can be used to reduce the size of candidate k-itemsets, Ck, for k > 1.
- Transaction reduction (reducing the number of transactions scanned In future iterations):A transaction that doesn't contain any frequent k-itemsets cannot contain any frequent (k+1) itemsets. Therefore, such a transaction can be marked or removed from further consideration because subsequent scans of database for j-itemsets where j > k, will not require it.
- Partitioning(Partitioning the data to find candidate itemsets):A partitioning technique can be used that requires just two database scans to mine the frequent itemsets. It consists of two phases. In phase I ,the algorithm sub divides the transaction of D into n overlapping partitions. If the minimum support threshold for transactions in D is min-sup, then the minimum support count for a partition is min-sup x the number of transactions in that partition. For each partition, all frequent itemsets within the partition are found. These are referred to as local frequent itemsets. In phase II, a second scan of D is conducted in which the actual support of each candidate is accessed in order to determine the global frequent itemsets.
- Sampling: The basic idea of the sampling is to pick a random sample S of given data D and then search for frequent itemsets in S instead of D.
- Dynamic itemset counting: In a dynamic itemset counting technique, new candidate itemsets can be partitioned into blocks by start points. The technique is dynamic in that it estimates the support of all of the itemsets that have been counted so far, adding new

candidate itemsets if all of their subsets are estimated to be frequent.

Algorithm: Apriori: Find frequent itemsets using an iterative level-wise approach based on candidate generation.

Input: D, a database of transactions; min-sup, the minimum support count threshold.
Output: L, frequent itemsets in D.

Method:

```
L1=find_frequent_1_itemsets(D);
for(k=2;Lk-1≠Ø;k++)
{
Ck = apriori_gen(Lk-1);
for each transaction t Є D
{
C1=subset(Ck,t);
for each candidate c Є Ct
c.count++;
}
Lk= {c Є Ck | c.count ≥ min_sup}
}
return L=UkLk;
```

Procedure apriori_gen (Lk-1 : frequent (k-1)-itemsets)

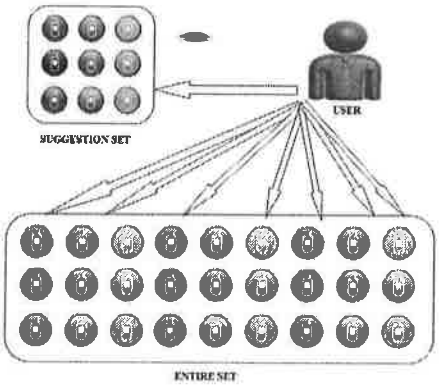
```
for each itemset l1 Є Lk-1
for each itemset l2 Є Lk-1
if(l1[1]=l2[1])^(l1[2]=l2[2])^.......^(l1[k-2]=l2[k-
2])^(l1[k-1]<l2[k-1]) then
{
c=l1∞l2;
if (has_infrequent_subsets(c,Lk-1) then
delete c;
else add c to Ck;
}
return Ck;
```

Procedure has_infrequent_subset (c:candidate k-itemset; **Lk-1:frequent(k-1) –itemsets);**

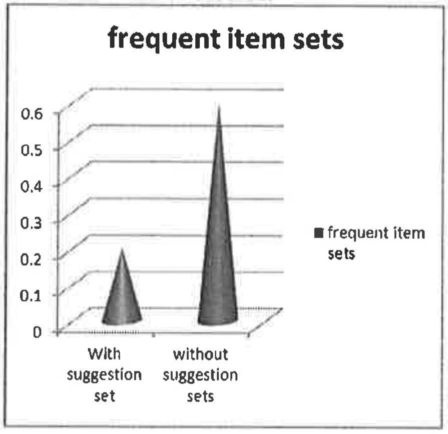
```
for each (k-1) –subset s of c
If s Є Lk-1 then
return TRUE;
return FALSE;
```

V.RESULTS

The above is the best method of ranking and suggesting the best methods in the scenario of mobile phone outlets in INDIA, which is shown in the following diagram:

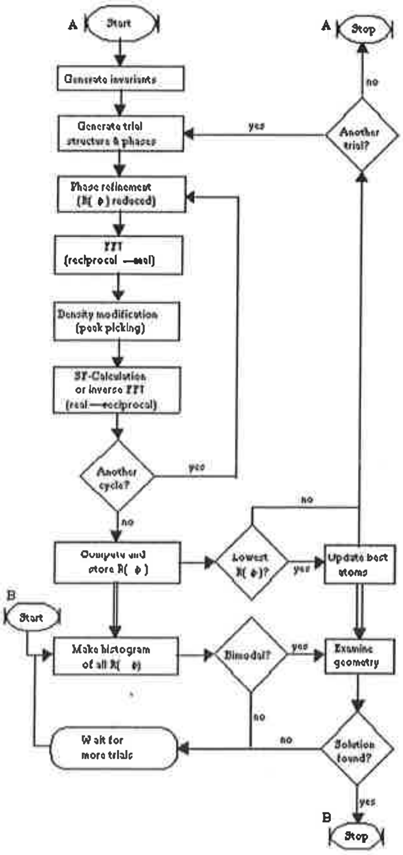


As it was shown in the above diagram we were going to take the most liked items from the users and suggesting the best mobiles or the best set of suggestions that the most of the users liked or ordered.



The confidence of the suggestions were also proved by an traditional confidence calculations as follows In this section we are going to discuss about algorithms. Till now we have discussed some ranking rules , suggestion rules and Frequency move2set algorithm. We have some problems with these, so we go for an algorithm which suits our requirements well. The algorithm is Apriori algorithm. In order to know these algorithms we need to know some concepts of data mining.

Data flow chart of proposed system:



Frequent itemsets: Let I={I1, I2, I3,…., Im} be a set of items. Let D, the task-relevant data, be a set of database transactions where each transaction T is a set of items such that T is a subset of I. Each transaction is associated with an identifier, called TID. Let A be a set of items. A transaction T is said to contain A if and only if A is a subset of T. An association rule is an implication of the form A > B, where A is subset of I, B is subset of I and A∩B =Ø. The rule A > B holds in the transaction set D with support s, where s is the percentage of transactions in D that contain AUB. This is taken to be the probability ,P(AUB).The rule A > B has confidence c in the transaction set D, where c is the percentage of transactions in D containing A that also contain B. This is taken to be the conditional probability, P(B/A). That is, Support(A=>B) = P(AUB) Confidence(A=>B) = P(B/A) Rules that satisfy both a minimum support threshold (min_sup) and a minimum confidence threshold (min_conf) are called strong. The occurrence frequency of an itemset is the number of transactions that contain the itemset. This is also known, simply as the frequency, support count,or count of the itemset. The set of frequent k-itemset is commonly denoted by Lk. confidence(A > B) = P (A / B) = support(AUB) / support(A) = supportcount(AUB) / supportcount(A).

Mining frequent itemsets: In general, association rule mining can be viewed as a two-step process: 1. Finding all frequent itemsets: By definition, each of these itemsets will occur at least as frequently as a predetermined minimum support count, min-sup. 2. Generate strong association rules from the frequent itemsets: By definition, these rules must satisfy minimum support and minimum confidence

VI.CONCLUSION

All the previous process already proposed were very complex and contains very complicated computations which made the ranking and suggesting the best and popular items have been more and more complex and not getting to the actual end users. Now we have proposed as very simple randomized algorithm for ranking and suggesting popular items designed to account for popularity bias. This was utilized by many of the mobile outlets in the country successfully.

REFERENCES

[1] Huidrom Romesh Chandra Singh, T. kalaikumaran, Dr. S. Karthik, Suggestion of True Popular Items, IJCSE, 2010.

[2] Y.Maanasa, V.Kumar, P.Satish Babu, Framework for suggesting POPULAR ITEMS to users by Analyzing Randomized Algorithms, IJCTA, 2011.

[3] V. Anantharam, P. Varaiya, and J. Walrand, —Asymptotically Efficient Allocation Rules for the Multiarmed Bandit Problem with Multiple Plays—Part i: i.i.d. Rewards,l IEEE Trans. Automatic Control, vol. 32, no. 11, pp. 968-976, Nov. 1987.

[4] J.R. Anderson, —The Adaptive Nature of Human Categorizationl Psychological Rev., vol. 98, no. 3, pp. 409-429, 1991.

[5] Yanbin Ye, Chia-Chu Chiang, A Parallel Apriori Algorithm for Frequent Itemsets Mining, IEEE, 2006.

[6] Cong-Rui Ji, Zhi-Hong Deng, Mining Frequent Ordered Patterns without Candidate Generation.

[7] Huang Chiung-Fen, Tsai Wen-Chih, Chen An-Pin, Application of new Apriori Algorithm MDNC to Exchange Traded Fund, International Conference on Computational Science and Engineering, 2009.

[8] Milan Vojnovi_c, James Cruise, Dinan Gunawardena, and Peter Marbach, Ranking and Suggesting Popular Items, IEEE, 2009.

IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 1, Ver. VII (Feb. 2014), PP 44-51
www.iosrjournals.org

Routing Framework for Delay Tolerant Networks using BayesiaLab

B.Sujatha Asst.professor, CH.Bhavani Asst.professor, A.Nagasri Asst.professor

Abstract: *Routing in delay tolerant networks (DTN) can benefit from the fact that most real life DTN, especially in the context of people-centric networks (e.g. Pocket Switching Networks (PSN)), exhibit some sort of periodicity in their mobility patterns. For example, public transportation networks follow periodic schedules. Even most individuals have fairly repetitive movement patterns, for example, driving to and from work at approximately the same time everyday. This paper proposes a BayesiaLab tool based DTN routing framework that adopts a methodical approach for computing the routing metrics by utilizing the network parameters (e.g. spatial and temporal information at the time of packet forwarding) that capture the periodic behavior of DTN nodes. After the calculation of routing metrics, different routing instantiations are possible based on this framework. We simulate a real-world vehicular DTN network using mobility traces from a metropolitan public transportation bus network and demonstrate that even a simplistic single-copy forwarding scheme based on our framework outperforms existing gradient-based single copy schemes by 25% in terms of delivery ratio. To the best of our knowledge this work is one of the first studies that adopts Bayesian inference in the context of DTN routing.*

I. Introduction

Delay Tolerant Networks (DTN) [3] are a class of challenged networks wherein the node contacts are intermittent and disconnections are common place. To deal with this episodic connectivity in DTN, the proposed routing strategies rely on the inherent mobility of the participating nodes to *store-carry-and-forward* [1] the messages for delivery to the destination. Such networks are well-suited in areas where there is no communication infrastructure due to harsh environments (battlefields, forests, space) or economic conditions (rural and remote areas, developing countries). Two particularly attractive instances of people-centric DTN include (i) *Pocket Switched Networks (PSN)* [7], wherein personal communication devices carried by humans self-organize to form an intermittently connected network, thus enabling a new class of social net- working applications (e.g. PeopleNet [15]) and (ii) Vehicular DTN, which can leverage the large data storage and energy capabilities offered by vehicles to create a large-scale powerful DTN. Examples include the use of vehicle-based DTN to provide low cost digital communication to remote villages [16] and vehicular sensing platforms such as CarTel [9] for urban monitoring.

Over the past years, several routing schemes have been proposed for DTN (an overview of these can be found in [11]).

A particularly effective forwarding principle that is often em- ployed is *gradient routing* [11], wherein the message tends to follow a gradient of increasing utility function values towards the destination. The utility function serves as the routing metric and is based on a variety of parameters, which depend on the *a priori* information about the network characteristics, such as: last encounter [2] with the destination, encounter frequencies of nodes [1], etc.

However, most of the schemes (e.g. [14], [1]) need to maintain routing metrics for all potential destinations, which is clearly not scalable, particularly for large-scale DTN (e.g. PSN or vehicle-based DTN). A more scalable approach, presented in [8], involves grouping of the nodes into a finite set of classes based on certain community affiliations (i.e. institutional affiliation) that exist amongst the nodes. Consequently, nodes only need to maintain routing metrics on a per-class basis. The goal is to transfer the message to any node that belongs to the same class as that of the destination since members of the same class have a high likelihood of encountering each other. However, this type of simple classification may be harder to achieve in larger and complex networks such as people-centric DTN because it is not obvious how and which properties should be utilized for the classification purpose.

In most people-centric DTN, the mobility patterns of the nodes generally exhibit some level of time periodicity. For example, public transport buses follow fixed schedules and routes; also individuals often follow repetitive patterns, for example - driving to and from work at the same time every weekday. Given that the mobility of most real-world DTN follow repetitive patterns to some extent, statistical approaches are particularly suitable for the above mentioned classification of nodes. In this paper, we present a novel routing framework, which utilizes the concept of Bayesian classification [6] for determining the class membership probabilities. The simplicity and accuracy of Bayesian classifier [6] make it an attractive candidate for this classification problem. In our approach the computation of these class membership probabilities relies on the availability of historical statistics about relevant network parameters (e.g. node mobility traces, node encounter statis- tics, etc.). After computing class membership probabilities, our routing framework employs class-based gradient routing similar to the ideas presented in [8]. The use of posterior probabilities in computing class membership enables us to accommodate more information as compared to prior probabilities.

For example, in vehicular DTN, packet forwarding behavior varies in different periods of a day due to heterogeneous traffic patterns throughout the day. Also landmarks like bus-stops are good place for packet forwarding since there is a high likelihood of finding a suitable forwarder in those areas. The effect of these additional attributes (e.g. time, location) can be factored in methodically by using posterior probabilities.

In particular, we make the following contributions:

- We introduce an extensible routing decision framework based on Bayesian classification, which seamlessly inte- grates knowledge about network characteristics and node mobility patterns in order to make better routing decision.
- We present a simple instantiation of our framework, referred to as *Bayesian*, which uses two simple classes based on prior packet delivery statistics. The routing metrics employed are posterior delivery probabilities conditioned on two attributes, location and time.
- We demonstrate the efficacy of our approach through simulation-based evaluations using mobility traces of a realworld public transport network. The simple *Bayesian* scheme achieves a 25% improvement in the packet delivery ratio as compared to that of MaxProp [1], an effective routing protocol that employs prior probabilities.

The rest of the paper is organized as follows. Section II provides an overview of related work. Section III presents our proposed Bayesian classifier based routing decision frame- work. A detailed example illustrating the use of this framework is also provided. Section IV presents the simulation-based evaluation of our scheme and compares it with other routing strategies. Finally, Section V concludes the paper.

II. Related Work

The idea of using prior information about network char- acteristics has been proposed in prior literature [1], [14], [2], [4], [11]. For example, the routing schemes proposed in [2] and [4] maintain age-of-last-encounter timers and choose a forwarding node that has most recently encountered the destination. In general, most of those routing schemes are based on the principle of Gradient Routing [11]. The basic idea is to transfer the message to the contact that has better delivery metric than the current incumbent. Lindgren et al. [13] introduced a probabilistic routing strategy wherein each node maintains a utility function, which is an exponential weighted moving average of prior contact probability for every other node. This utility function is then used as a metric for gradient routing. Burgess et al. [1] use incremental averaging of node encounters to calculate the delivery predictability.

However, most of the above mentioned stochastic routing protocols use prior probabilities (i.e. the probability of an event regardless of other events) in making a routing decision. For example, MaxProp [1] uses unconditional encounter-based prior probabilities to calculate shortest path. On the contrary, our proposed decision model is based on posterior probabilities, i.e., the probability of an event when relevant other attributes are taken into account. Consequently, this enables our scheme to make a more informed decision in choosing suitable forwarders and thus improves the routing performance in terms of delivery ratio.

In machine learning research, studies [12] comparing classification algorithms have found that even a naive Bayesian classifier performs comparably with Logistic Regression and Support Vector Machine . Text classification, weather predition, large database management (e.g. in NASA's space flight centre) are some of the practical uses of Bayesian classifiers. To the best of our knowledge this is one of first attempts at adopting bayesian inference in DTN routing.

BayesiaLab

A single tool for all your Decision Support and Data Mining Tasks

- Bayesian networks will allow you to model your expert knowledge relative to risk, fraud, customer s' behavior ...The graphical representation of Bayesian networks and the BayesiaLab's ease of use make it a invaluable Brain Storming and communication tool.
- You will be able to exploit all the power of unsupervised learning to extract from your data bases the set of probabilistic relations that are really significant. This kind of learning is a real knowledge discovery tool and is very helpful for the understanding of your problems.
- Supervised learning will allow you to characterize your target variable. This variable will represent, for example, the fraud, the propensity to buy a product, or the customer satisfaction. The evaluation of the automatically learnt Bayesian network on an independent test set (cases that have not been used for learning) will return you the model global precision, its confusion matrix (occurrences, reliability and precision) that will

enable you to precisely know the prediction behavior of the network. You will also have an interactive lift curve that will help you finding the threshold representing the best economic compromise for your marketing actions.
• You will also be able to exploit the powerful Markov Blanket Learning algorithm for the selection of the minimal subset of variables that are really useful with respect to the target variable.
• Clustering of your data bases will enable you to discover groups of homogeneous individuals sharing the same characteristics. The HTML report that will give you the probabilistic profile of each identified cluster will help your experts to put a name to these clusters and to use them in your marketing campaigns.
• If Expert knowledge is available, BayesiaLab will rigorously merge it with your data bases.
• BayesiaLab will enable you testing levers effects (e.g. action for image improvement or for training) by manually adding nodes to your learnt networks. By associating cost nodes to these levers, you then will be able to evaluate various action policies.
• The BayesiaLab's adaptive questionnaires will return you the most relevant questions with respect to the information brought to the knowledge of your target variable and with respect to the cost associated to questions. A new set of ordered questions will be automatically returned after each answer.
• You will be able to use your Bayesian networks off-line to automatically classify new cases described in a data base. Two additional fields will be added to each case: the predicted value and its probability.
• By using the BayesiaLab's analysis toolbox, you will really be able to understand your data: analysis of the strength of the relations, analysis of the interaction between your target variable and the other variables, analysis of the relations linking all the variables with a specific value of the target variable, contradiction analysis to know if all the evidences support the same conclusion or if there are some contradicting evidences, causal analysis to transform the arcs that can be inverted without changing the probabilistic meaning of the network into edges.
• You will be able to "play" with your networks to easily test your hypothesis by carrying out What-if scenarios.
• Lastly, you will have access to a rigorous imputation tool that will use your Bayesian network jointly with all the available evidences to compute the probability distribution of your missing values, and then Replace them accordingly.

III. Bay E S I A N Routing Framewo Rk

Our routing framework consists of two phases: 1) classification and 2) packet forwarding. The goal of the first phase is to create an abstract grouping of nodes based on certain prior information about the network. Gradient routing is employed in the second phase, wherein a node tries to forward the packet to a neighbor which has higher *affinity* with the destination, in anticipation that the packet gradually moves towards destination. The affinity of a node with the destination is computed as a posterior probability using Bayesian inference and is based on the history of past packet forwarding behavior which is conditioned on network specific factors such as: node location, time etc. The intuition being that a suitable forwarder which was encountered at the same time of day and location in the past is likely to be a good candidate in the future. Bayesian inference allows us to model the imprecise node mobility and outcomes of interest (e.g. routing success or failure) by combining common-sense knowledge and observational evidence. Its ability to express all forms of uncertainty in terms of probability makes it an attractive tool to quantify a node's probability (i.e. affinity) to meet destination.

The proposed routing framework has two phases which we elaborate in the next section, followed by an instantiation of the framework. It should be noted that depending on the network attributes (e.g. node location, forwarding time, etc.), several routing instantiations are possible based on this framework.

A. Phase One: Bayesian classification

Generally speaking, a Bayesian classifier takes the attributes of an unknown sample and tries to predict its class membership probability based on the history of previously known samples. Let us assume, a forwarding node P has
${}^{x}N_1, N_2, \ldots, N_m$ neighbors which belong to one of the classes ${}^{x}C_1, C_2, \ldots, C_n$, where m n. In practice, a node may belong to more than one classes but its class membership probabilities would usually be different. Now, assume that node P has a packet to send to destination D. Node P and node D are members of classes C_P and C_D, respectively. If node P encounters another node that has a higher affinity towards the destination class C_D, then P should pass on the message to this node.
Let, the attributes (which have direct or indirect effect on the packet delivery) of a node be represented by an attribute vector, $X = {}^{x}x_1, x_2, \ldots, x_n$. Current time, physical location,
contact probability and contact duration with destination are some examples of attributes, which affect packet delivery probability. It is upto the network administrator (or system developer) to determine appropriate classes and attributes for a particular network.

Now, referring back to our example, a node can calculate its affiliation probability with other classes (C_i) conditioned on the attributes X (i.e. posterior probability) as: $P(C_i|X)$. Intuitively, a larger value of the posterior probability would imply better affiliation to that class. The posterior probability $P(C_i|X)$ is based on more information than prior probability $P(C_i)$ because it factors in the effect of the attributes on the class memberships, and is hence able to identify better forwarding candidates. Using Bayes theorem, the posterior probability can be calculated from prior probability as:

$P(C_i|X) = P(X|C_i)P(C_i)/P(X)$ (1) The denominator of Eq. 1, $P(X)$ does not depend on C_i and is only used for normalization. Clearly, $P(C_i|X)$ will be maximum when $P(X|C_i)P(C_i)$ is maximum. To quantify the affiliation of a node to a class, we define a metric called the *Affiliation Index*, which is computed using Eq. 2.

$$\text{Affiliation Index} = P(X|C_i)P(C_i) \qquad (2)$$

$P(X|C_i)$ and $P(C_i)$ can be calculated based on the available historical information.

The term $P(X|C_i)P(C_i)$ is equivalent to the joint prob- ability model of $P(C_i \cap X)$. However, an expansion of $P(X|C_i)P(C_i)$ becomes unmanageable for dimension of $X =$ $^x x_1, x_2, \ldots, x_n$. If we can assume that every attribute x_i is conditionally independent on every other attribute x_j $(i = j)$ then we can come up with Eq. 3.

$$P(X|C_i) = \quad \overset{n}{\underset{k=1}{}} P(x_k|C_i) \qquad (3)$$

Despite the naive assumption of conditional independence among the attributes, the naive Bayesian classifier exhibits remarkable accuracy in classification process [17]. Now, the affiliation index can be computed from Eq. 2 and 3 as:

$$\text{Affiliation Index} = P(C_i) \underset{\forall k}{} \quad P(x_k|C_i) \qquad (4)$$

The conditional probability $P(x_k|C_i)$ can be estimated from the history data set. Assume that the historical data contains s_i samples that belong to class C_i. Assume that from this sample set, there are s_{ik} samples for which the value of an attribute is equal to x_k. In this case, $P(x_k|C_i) = {}^{s_{ik}}$. If it is known that certain attributes follow well-known probability distributions (e.g. exponential, gaussian, etc.), then we can directly usethat distribution in place of$_s P(x_k|C_i)$. It should be noted that if any of the $P(x_k|C_i)$ in Eq. 4 becomes zero, then the corresponding affiliation index will also be zero, nullifying the effect of other attributes. In order to circumvent this problem, we assume a low value (e.g. 0.01) for $P(x_k|C_i)$ whenever its value becomes zero.

B. Phase Two: Forwarding

The second phase of our framework uses the affiliation indices (computed in the first phase) to forward the packets towards the destination in gradient manner. When a node encounters one or more neighbors, it needs to make a decision on which node is a suitable forwarder. The node can readily compute its own affiliation index with the destination using Eq. 4. However, it also needs to know the affiliation indices of its neighbors with the destination. The node can calculate the class affiliation probabilities for all other neighbors provided that the node knows the attributes of its neighbors. However, calculating the posterior class affiliation probabilities is com- putationally non-trivial. This is especially true if there are a large number of neighbors (i.e. dense topology) and/or if there are a large numbers of classes and attributes to be considered for calculating the posterior probabilities. For example, in vehicular DTN, some regions (e.g. bus stops, road crossings etc.) have dense vehicle distributions and hence the number of neighbors are usually much large in these areas. It should also be noted that these computations need to be performed on per- packet basis. In order to reduce the computation overhead, the forwarding node distributes the calculation of the affiliation indices among its neighbors using a simple query-response based protocol: the forwarder broadcasts a request containing its own affiliation index with the destination and overhearing neighbors respond it back with their own affiliation indices. As a result, the forwarder can select the neighbor which has maximum affinity towards destination.

In the next sub-section, with the help of an illustrative example, we discuss a simplified single-copy version of our scheme. A multi-copy version can also be implemented in a straight forward manner.

C. Example

We present a simplified example to illustrate the operation of our routing protocol. We consider the context of a vehicle- based DTN. It is assumed that the past delivery statistics are already known. In public transportation networks, where nodes are equipped with huge storage, computing power and sensors, these information together with nodes' coordinate can be logged as the nodes deliver packets to destinations. We assume a very basic classification scheme consisting of
only two classes: *delivered*, which includes nodes which acted as relays in successfully delivering packets to the destination, and *not-delivered* which indicates nodes which failed to deliver packets. For simplicity, we have chosen two attributes ${}^{x}x_1, x_2$ which are:

• x_1 = *Region Code* ${}^{x}R_1, R_2, \ldots, R_m$, identifying the location where the node was situated at the time of packet forwarding. We assume that the entire physical domain of the network is divided into m smaller rectangular grids of equal size, each with a distinct identifier. The motivation for using location as an attribute is because in most DTN, the packet delivery probability depends on the physical location of the nodes. For example, in vehicular DTN, the probability of finding a suitable forwarder is higherin the vicinity of landmarks like bus-stops and parking lots, than on isolated suburban roads.

• x_2 = *Time Slot* ${}^{x}T_1, T_2, \ldots, T_n$, indicating the time slot when the packet was forwarded. This metric has been chosen because packet delivery also depends on different time periods of a day. In vehicle-based DTN, different traffic patterns are prevalent in different periods of a day and hence the packet forwarding behavior also varies with time. The granularity of the time slot as well as region code can be increased or decreased depending on the availability of historical data. In general, choosing finer grained time slots will result in a more precise affiliation index but requires larger historical records.

In the rest of this section, we explain the operations of our proposed scheme in the context of a vehicular DTN.

A sample topology of a vehicular DTN and past packet forwarding statistics are shown in Fig. 1. The vehicles repre- sent nodes and the links represent the connectivity among the neighbor nodes. Each node maintains a history of past packets that have been forwarded. The history can be gathered via the the propagation of acknowledgments (details in Section IV-C). Each tuple of the history database contains: (a) destination address of the packet, (b) region code where the node was situated at the time of packet forwarding, (c) time slot when the packet was forwarded and (d) the class of the packet, i.e., whether the packet was successfully delivered to the destination or not). We use a simple YES and NO to indicate class memberships. For example, the first entry of the history data maintained by node A (${}^{x}D, R_1, T_2$, YES) indicates that while node A was in region R_1, it had forwarded a packet (with destination D) at time slot T_2 and the packet eventually reached the destination.

Let us assume that node A has a packet to send to node
D; A is in region R_2 and the current time slot is T_1. Node A calculates its *affiliation index* for D using Eq. 2, which is elaborated below. The prior class probability (for class $C_{delivered}$) is: $P\ (C_{delivered}) = 0.5$ (since there are two entries in the training dataset of node A for destination D, from which one tuple belongs to the class labeled *yes* or next hop. This process continues until the packet reaches the destination or is discarded (due to expired TTL value).

IV. Simulation-Based EvalUatIONs

To demonstrate the efficacy of our approach we have carried out a simulation-based evaluation in the context of a vehicular DTN. We consider a metropolitan public transport bus network [10] and assume that each bus is equipped with a wireless radio, thus simulating a large-scale DTN. We choose to use a public transport network because of the inherent repetitive nature of the nodes movements.

We use the simple example discussed in Section III-C (with two classes *delivered* and *non-delivered* and two attributes: *region code* and *time slot*) as an instantiation of our framework. This is referred to as *Bayesian* in the rest of the simulations. We compare this scheme with three other routing strategies: (i) *Epidemic* [18], (ii) single copy version of *MaxProp* [1] and (iii) *Wait* (or direct delivery) [11]. Epidemic and Wait exhibit greatly contrasting properties. Epidemic routing is known to achieve the best case delivery ratio with minimum delay since it relies on flooding. On the contrary, Wait requires the minimum possible cost to deliver a packet since the packet is directly forwarded to the destination without the need for relays. In this study, we assume that there are sufficient buffers at each node, since we wish to exclusively focus on the performance of the routing strategies. The simulations have been conducted using a custom built simulator and PostgreSQL has been used as the back-end database server for storing and processing the history data set at each node.

A. Details of Mobility Traces

We have used the mobility traces of buses in the King County Metro bus system in Seattle, USA [10] to simulate a DTN network. This public-transport system consists of 1163 buses plying over 236 distinct bus routes covering an area of 5100 square kilometers. The traces were collected over a two week period in November 2001. The traces are based on location update messages sent by each bus.

Our *Bayesian* protocol requires$_2$information about the loca-*delivered*; hence the probability is: ^{1}or 0.5). Now, in order tion (region code) and time (time slot). For this purpose, we to calculate $P\ (X|C_i)$ where $x_1\ =\ R_1$, $x_2\ =\ T_1$, we need to calculate: $P\ (x_1|C_{delivered})$, $P\ (x_2|C_{delivered})$. From the history data of A (Fig. 1), we find, $P\ (x_1|C_{delivered})$ = 0.01 and $P\ (x_2|C_{delivered})$ = 0.01. Recall that null values need to be replaced by a small representative value (0.01 in this example). Now the node can calculate the *affiliation index*, I, using, Eq. 2 as, assume that the entire region under consideration is divided into 1km × 1km square grid blocks. In general, selection of smaller grid size results in a more precise affiliation index but increases the amount of historical data that needs to be maintained to calculate statistically significant posterior probabilities. For similar reasons, the granularity of the time slots is chosen as 10 minutes (600 seconds).

$I = P\,(C_{delivered})P\ (x_1\,|C_{delivered})P\ (x_2\,|C_{delivered})$

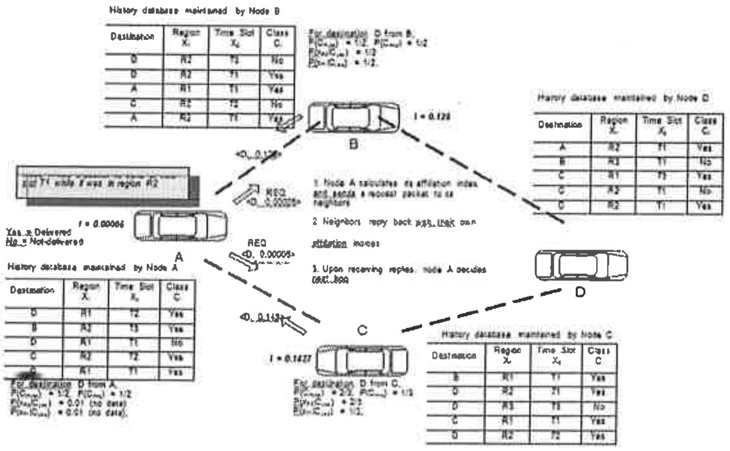
or, $I = 0.5 \times 0.01 \times 0.01 = 0.00005$

Node A broadcasts a request packet containing destination node and its delivery index ${}^{x}D, I$ to its neighbors. Upon re- ceiving the request, the neighbors B and C also calculate their own affiliation index, which are 0.125 and 0.1437 respectively. Since these are greater than that of node A, both nodes transmit their respective indices to A. Node A then chooses C as the

Though we have mobility traces for the entire 24 hour duration of a weekday, we focus on a 9 hour period from 6am -

3pm to ensure that the simulations are tractable. This period is sufficient to capture the periodicity of the bus mobility, since a typical trip along a bus route takes 2 - 2.5 hours. For the same reasons, we have concentrated on a 58 km x 88 km region, which includes the central business district of the city. Further,

History database maintained by Node B


History database maintained by Node B
For destination D from B,
P(C_del) = 1/2, P(C_nd) = 1/2
P(x1|C_del) = 1/2
P(x2|C_del) = 1/2.
I = 0.125
B
<D, 0.125>
History database maintained by Node D
slot T1 while it was in region R2
REQ <D, 0.00005>
1. Node A calculates its affiliation index and sends a request packet to its neighbors
2. Neighbors reply back with their own affiliation indices
3. Upon receiving replies, node A decides next hop
I = 0.00005
Yes = Delivered
No = Not-delivered
A
REQ <D, 0.00005>
D
History database maintained by Node A
<D, 0.1437>
C
History database maintained by Node C
I = 0.1437
For destination D from A,
P(C_del) = 1/2, P(C_nd) = 1/2
P(x1|C_del) = 0.01 (no data)
P(x2|C_del) = 0.01 (no data).
For destination D from C,
P(C_del) = 2/3, P(C_nd) = 1/3
P(x1|C_del) = 2/3
P(x2|C_del) = 1/2.

History database maintained by Node B

Destination	Region X_1	Time Slot X_2	Class C_i
D	R2	T2	No
D	R2	T1	Yes
A	R1	T1	Yes
C	R2	T2	No
A	R2	T1	Yes

History database maintained by Node D

Destination	Region X_1	Time Slot X_2	Class C_i
A	R2	T1	Yes
B	R3	T1	No
C	R1	T3	Yes
C	R3	T1	No
D	R2	T1	Yes

History database maintained by Node A

Destination	Region X_1	Time Slot X_2	Class C
D	R1	T2	Yes
B	R3	T3	Yes
D	R1	T1	No
C	R2	T2	Yes
C	R1	T1	Yes

History database maintained by Node C

Destination	Region X_1	Time Slot X_2	Class C
B	R1	T1	Yes
D	R2	T1	Yes
D	R3	T3	No
C	R1	T1	Yes
D	R2	T2	Yes

Fig. 1. Illustrative Example of our Routing Framework

we only consider the 35 buses whose routes are for the most part contained within the region of interest.

We use 50 random source-destination pairs (from the combination of 35 × 35 unique source-destination pairs) where each source is a CBR node which transmits a new packet of size 1000 bytes every 100 seconds. The source nodes begin to transmit packets at time $1000+t_{rand}$ ($0 < t_{rand} < 4600$) after the simulation starts in order to ensure that all the buses in our trace file become active by this time. The whole simulation runs for 32400 seconds (i.e. 9 hours). The entire simulation is repeated 20 times with different sets of source-destination pairs and the results presented are averaged over these runs.

C. Gathering Packet Traversal History

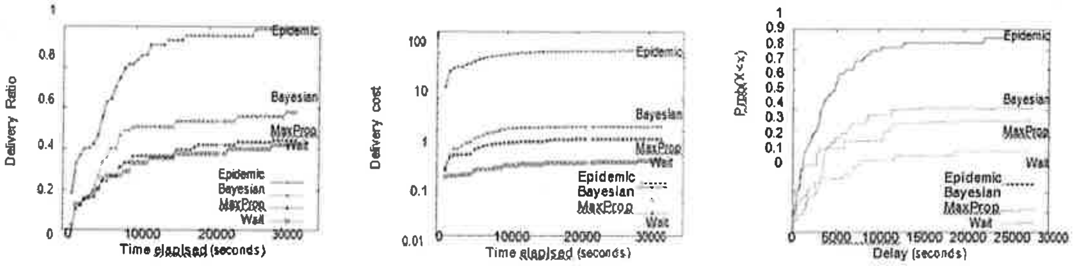
Recall that our routing framework requires prior statistics about certain network parameters, which can serve as the input to the Bayesian classifier. In particular, the example *Bayesian* protocol requires past packet delivery statistics and information about the region code and time slot when these packets were forwarded. In order to emulate the requisite history, we simulated 500 sample runs of the simulation with exactly the same parameters as discussed in Section IV-B. The only difference is that we used a 9 hour period (6am - 3pm) for the weekday prior to that used for the actual evaluations. Though we have not updated the history in our simulation (since lack of updated history only affects long-lasting simulations), it can be easily done in real-world scenario. In our instantiation, the nodes need to know if the forwarded packets eventually reach the destination or not. Nodes which deliver the packets to the destination can readily log these packet traversal history. Or, some form of acknowledgment packets can be used to let the forwarders to update their records. In fact, in many multi-copy DTN routing protocols, acknowledgment packets (e.g. *vaccine* [5]) are sent back to the sender so that the sender and intermediate nodes can purge the packet from their queues. These acknowledgments can be used to update the history.

D. Delivery Ratio Results

The most important metric in DTN routing is the delivery ratio, which is the ratio of the messages delivered to the mes- sages created. As seen from Fig. 2a, the delivery ratio achieved by Epidemic routing is near perfect, which is expected. Wait performs the poorest since it relies on direct delivery. Bayesian outperforms MaxProp by 25%. The reason behind that is, MaxProp suffers from inferior quality encounter probabilities (due to its prior probability calculation method which does not depend on other network parameters) and hence often makes non-optimal routing decisions. It should be noted that MaxProp performs path lookup upto *n* hops away to find the best route to the destination. This method of searching all possible paths is clearly not scalable in large networks. On the contrary, our framework computes the posterior probabilities in a distributed manner and hence can scale easily.

E. Delivery Cost Results

We measure the delivery cost by counting the total number of messages transmitted and normalizing it by the total number of unique message created. The Epidemic routing protocol achieves its higher delivery ratio with a very high cost. Fig.
2b shows the semi-log plot of delivery cost vs. time. The delivery cost is about 30 times less in our method compared to that of Epidemic routing. The delivery costs of MaxProp and our proposed method are similar, though Bayesian exhibits marginal better performance. This is because they both are



(a) Comparison of Delivery Ratios

(b) Comparison of Delivery Costs

(c) Cumulative Delay Distributions

Fig. 2. Simulation Results

single copy schemes and the delivery costs have been cal- culated from the delivered packets only, neglecting the fact that the schemes have different delivery ratios. So the single- copy routing methods which have low delivery ratios have low delivery costs in general. Besides, we have not considered the added cost of exchanging the summary vectors which is an integral part of MaxProp. The need for exchanging summary vectors also limits MaxProp's deployment in large networks.

F Delay Distribution Results

Though a large packet delay is common in delay tolerant networks, it is always desirable to receive a packet as early as possible. In order to get an idea of packet delivery delay, we plot the cumulative delay distribution of the routing protocols in Fig. 2c. Epidemic routing achieves the minimum possible packet delivery delay (receives 50% of the packets during first≈3000 seconds) whereas the direct delivery approach (Wait) shows worst case delivery delay (receives only 20% packets during the first 3000 seconds). Bayesian and MaxProp again exhibit similar properties with Bayesian having a slight edge. It can be concluded from the results above that the deliv- ery ratio of even a simple implementation (which uses two simplistic classes and just two attributes) of our framework (which is a single-copy method) is about 60% of the best case (i.e. multi-copy Epidemic), while incurring an overhead that is 30 times less than that of the base case. The delivery ratio is significantly better than that of MaxProp. Also, it is intuitive that the use of better attributes further improves the performance of routing protocol.

V. Conclusion

In this paper, we present a Bayesian classifier based routing decision framework which simplifies the integration of various routing attributes and utilizes the repetitive nature of people- centric DTN in making better routing decisions. Preliminary simulation study with traces from a real-world delay tolerant public transport network demonstrates considerable perfor- mance gain (about 25% improvement in terms of packet delivery ratio) than existing gradient-based routing schemes.

It is intuitive that choosing proper attributes and classes in our routing framework may increase the performance of the routing protocol even further. We intend to investigate methodologies for choosing appropriate attributes and classes for our routing framework. We also plan to evaluate the effectiveness of our framework in other types of DTN with imprecise schedules (e.g. PSN, etc.).

References

[1] J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine. Maxprop: Routing for vehicle-based disruption tolerant networks. In *Proceedings of IEEE INFOCOM*, April 2006.

[2] H. Dubois-Ferriere, M. Grossglauser, and M. Vetterli. Age matters: Efficient route discovery in mobile ad hoc networks using encounter ages. In *Proceedings of ACM MobiHoc*, June 2003.

[3] K. Fall. A delay-tolerant network architecture for challenged internets. In *Intel Research Technical Report, IRB-TR-03-003*, February 2003.

[4] M. Grossglauser and M. Vetterli. Locating nodes with ease: Last encounter routing in ad hoc networks through mobility diffusion. In *Proceedings of IEEE INFOCOM*, March 2003.

[5] Z. J. Haas and T. Small. A new networking model for biological applications of ad hoc sensor networks. In *ACM Transaction on Networking*, volume 14, 2006.

[6] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, ISBN: 1558609016, 2001.

[7] P. Hui, A. Chaintreau, J. Scott, R. Gass, C. Diot, and J. Crowcroft. Pocket switched networks and human mobility in conference environ- ments. In *Proceedings of SIGCOMM*, August 2005.

[8] P. Hui and J. Crowcroft. How small labels create big improvements. In *Proceedings of International Workshop on Intermittently Connected Mobile Ad hoc Networks*, March 2007.

[9] B. Hull et al. Cartel: A distributed mobile sensor computing platform.In *Proceedings of ACM SenSys*, October 2006.

[10] J. G. Jetcheva, Y. C. Hu, S. P. Chaudhuri, A. K. Saha, and D. B. Johnson. Design and evaluation of a metropolitan area multitier wireless ad hoc network architecture. In *Proceedings of fifth IEEE workshop on Mobile Computing Systems and Applications*, October 2003.

[11] E. P. C. Jones and P. A. S. Ward. Routing strategies for delay-tolerant networks. In *Proceedings of ACM SIGCOMM*, August 2004.

[12] P. Langley, W. Iba, and K. Thompson. An analysis of bayesian classifiers. In *Proceedings of the 10th National conference on Artificial Intelligence*, July 1992.

[13] A. Lindgren, A. Doria, and O. Schelen.Probabilistic routing in intermittently connected networks. In *Proceedings of ACM MobiHoc*, June 2003.

[14] S. Merugu, M. Ammar, and E. Zegura. Routing in space and time in networks with predictable mobility. Technical report, Georgia Institute of Technology Technical report, GIT-CC-04-07.

[15] M. Notani, V. Srinivasan, and P. Nuggenhalli. Peoplenet: Engineering a wireless virtual social network. In *Proceedings of ACM MobiCom*, October 2004.

[16] A. Pentland, R. Fletcher, and A. Hasson. Daknet: Rethinking connec- tivity in developing nations. In *IEEE Computer*, volume 37(1), January 2004.

[17] I. Rish. An empirical study of the naive bayes classifier. In *IJCAI Workshop on Empirical Methods in Artificial Intelligence*, August 2001.

[18] A. Vahdat and D. Becker. Epidemic routing for partially connected ad hoc networks. In *Technical Report CS-200006, Duke University*, 2000.

IOSR IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 2, Ver. II (Mar-Apr. 2014), PP 40-46
www.iosrjournals.org

Efficient Detection of Internet Worms Using Data Mining Techniques

B. Sujatha, G.Rajitha devi
Asst.professor, lecturer

Abstract: *Internet worms pose a serious threat to computer security.Traditional approaches using signatures to detect worms pose little danger to the zero day attacks. The focus of malware research is shifting from using signature patterns to identifying the malicious behavior displayed by the malwaresThis paper presents a novel idea of extracting variable length instruction sequences that can identify worms from clean programs using data mining techniques.The analysis is facilitated by the program control flow information contained in the instruction sequences. Based upon general statistics gathered from these instruction sequences we formulated the problem as a binary classification problem and built tree based classifiers including C5.0, boosting and random forest. Our approach showed 99.5% detection rate on novel worms whose data was not used in the model building process.*

Keywords
- *Data Mining,*
- *Worm Detection,*
- *C5.0,Boosting,*
- *Feture selection using stepwise forward selection,*
- *static analysis,*
- *Diassembly,*
- *Insruction sequences.*

I. Introduction

Computer virus detection has evolved into malware detection since Cohen first formalized the term computer virus in 1983 [13]. Malicious programs, commonly termed as malwares, can be classified into virus, worms, trojans, spywares, adwares and a variety of other classes and subclasses that sometimes overlap and blur the boundaries among these groups [24]. The most common detection method is the signature based detection that makes the core of every commercial anti-virus program. To avoid detection by the traditional signature based algorithms, a number of stealth techniques have been developed by the malware writers. The inability of traditional signature based detection approaches to catch these new breed of malwares has shifted the focus of malware research to find more generalized and scalable features that can identify malicious behavior as a process instead of a single static signature. The analysis can roughly be divided into static and dynamicanalysis. In the static analysis the code of the program is examined without actually running the program while in dynamic analysis the program is executed in a real or virtual environment. The static analysis, while free from the execution overhead, has its limitation when there is a dynamic decision point in the programs control flow. Dynamic analysis monitors the execution of program to identify behavior that might be deemed malicious. These two approaches are combined also [23] where dynamic analysis is applied only at the decision-making points in the program control flow. In this paper we present a static analysis method using data mining techniques to automatically extract behavior from worms and clean programs. We introduce the idea of using sequence of instructions extracted from the disassembly of worms and clean programs as the primary classification feature. Unlike fixed length instructions or n-grams, the variable length instructions inherently capture the programs control flow information as each sequence reflects a control flow block. The difference among our approach and other static analysis approaches mentioned in the related research section are as follows. First, the proposed approach applied data mining as a complete process from data preparation to model building. Although data preparation is a very important step in a data mining process, almost all existing static analysis techniques mentioned in the related research section did not discuss this step in detail except [25]. Second, all features were sequences of instructions extracted by the disassembly instead of using fixed length of bytes such as n-gram. The advantages are:

- The instruction sequences include program control flow information, not present in n-grams.
- The instruction sequences capture information from the program at a semantic level rather than syntactic level.
- These instruction sequences can be traced back to their original location in the program for further analysis

of their associated operations.

- These features can be grouped together to form additional derived features to increase classification accuracy.
- A significant number of sequences that appeared in only clean program or worms can be eliminated to speed up the modeling process.
- The classifier obtained can achieve 95% detection rate for new and unseen worms.
- It is worth noting that a dataset prepared for a neural network classifier might not be suitable for other data mining techniques such as decision tree or random forest.
-

II. Related Research

[18] Divided worm detection into three main categories; Traffic monitoring, honey pots and signature detection. Traffic analysis includes monitoring network traffic for anomalies like sudden increase in traffic volume or change in traffic pattern for some hosts etc. Honeypots are dedicated systems installed in the network to collect data that is passively analyzed for potential malicious activities. Signature detection is the most common method of worm detection where network traffic logs, system logs or files are searched for worm signatures.

Data mining has been the focus of many malware researchers in the recent years to detect unknown malwares.

A number of classifiers have been built and shown to have very high accuracy rates. Data mining provides the means for analysis and detection of malwares for the categories defined above. Most of these classifiers use n-gram or API calls as their primary feature. An n-gram is a sequence of bytes of a given length extracted from the hexadecimal dump of the file. Besides file dumps, network traffic data and honey pot data is mined for malicious activities. [17] introduced the idea of using tell-tale signs to use general program patterns instead of specific signatures. The tell-tale signs reflect specific program behaviors and actions that identify a malicious activity. Though a telltale sign like a sequence of specific function calls seems a promising identifier, yet they did not provide any experimental results for unknown malicious programs. The idea of tell-tale signs was furthered by [10] and they included program control and data flow graphs in the analysis. Based upon the tell-tale signs idea, they defined a security policy using a security automata. The flow graphs are subjected to these security automata to verify against any malicious activity. The method is applied to only one malicious program. No other experimental results were reported to describe algorithm efficiency, especially on unseen data. In another data mining approach, [20] used three different types of features and a variety of classifiers to detect malicious programs. Their primary dataset contained 3265 malicious and 1001 clean programs. They applied RIPPER (a rule based system) to the DLL dataset. Strings data was used to fit a Naive Bayes classifier while n-grams were used to train a Multi-Naive Bayes classifier with a voting strategy. No n-gram reduction algorithm was reported to be used. Instead data set partitioning was used and 6 Naive-Bayes classifiers were trained on each partition of the data. They used different features to built different classifiers that do not pose a fair comparison among the classifiers. Naive-Bayes using strings gave the best accuracy in their model. A similar approach was used by [15], where they built different classifiers including Instance-based Learner, TFIDF, Naive-Bayes, Support vector machines, Decision tree, boosted Naive-Bayes, SVMs and boosted decision tree. Their primary dataset consisted of 1971 clean and 1651 malicious programs. Information gain was used to choose top 500 n-grams as features. Best efficiency was reported using the boosted decision tree J48 algorithm. [9] used n-grams to build class profiles using KNN algorithm. Their dataset was small with 25 malicious and 40 benign programs. As the dataset is relatively small, no ngram reduction was reported. They reported 98% accuracy rate on a three-fold cross validation experiment. It would be interesting to see how the algorithm scale as a bigger dataset is used. [22] proposed a signature based method called SAVE (Static Analysis of Vicious Executables) that used behavioral signatures indicating malicious activity. The signatures were represented in the form of API calls and Euclidean distance was used to compare these signatures with sequence of API calls from programs under inspection. Besides data mining, other popular methods includes activity

Monitoring and file scanning. [19] proposed a system to detect scanning worms using the premises that scanning worms tend to reside on hosts with low successful connections rates. Each unsuccessful or successful connection attempt was assigned a score that signals a host to be infected if past a threshold. [14] proposed behavioral signatures to detect worms in network traffic data. [16] developed Honeycomb, that used honeypots to generate network signatures to detect worms. oneycomb used anomalies in the traffic data to generate signatures. All of this work stated above, that does not include data mining as a process, used very few samples to validate their techniques. The security policies needed human experts to devise general characteristics of malicious programs.

Data preparation is a very important step in a data mining process. Except [25], none of the authors presented above have discussed their dataset in detail. Malicious programs used by these researchers are very eclectic in nature exhibiting different program structures and applying the same classifier to every program does not guarantee similar results.

III. Data Processing

Our collection of worms and clean programs consisted of 2775 Windows PE files, in which 1444 were worms and the 1330 were clean programs. The clean programs were obtained from a PC running Windows XP. These include small Windows applications such as calc, notepad, etc and other application programs running on the machine. The worms were downloaded from [8]. The dataset was thus consisted of a wide range of programs, created using different compilers and resulting in a sample set of uniform representation. Figure 3 displays the data processing steps.

❖ Malware Analysis

We ran PEiD [5] and ExEinfo PE [2] on our data collection to detect compilers, common packers and cryptors, used to compile and/or modify the programs. Table 1 displays

Table 1. Packers/Compilers Analysis of Worms

Packer/Compiler	Number of Worms
ASPack	77
Borland	110
FSG	31
Microsoft	336
Other Not Packed	234
Other Packed	83
PECompact	26
Unidentified	140
UPX	67

Table 2. Packers/Compilers Analysis of Worms and Clean Programs

Type of Program	Not Packed	Packed	Unidentified
Clean	1002	0	49
Worm	624	340	140
Total	1626	340	189

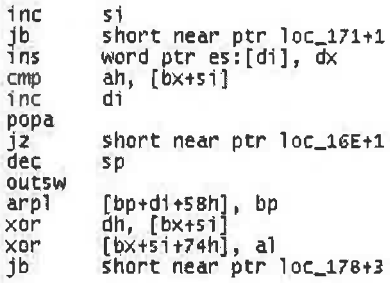
The distribution of different packers and compilers on the worm collection. The clean programs in our collection were also subjected to PEiD and Exelnfo PE to gather potential packers / crytpors information. No packed programs were detected in the clean collection. Table 2 displays the number of packed, not packed and unidentified worms and clean programs. Before further processing, packed worms were unpacked using specific unpackers such as UPX (with -d switch) [6], and generic unpackers such as Generic Unpacker Win32 [3] and VMUnpacker [7].

❖ File Size Analysis

Before disassembling the programs to extract instruction sequences, a file size analysis was performed to ensure that the number of instructions extracted from clean programs and worms is approximately equal. Table 3 displays the file size statistics for worms and clean programs. Table 3 indicates the that the average size of the clean programs is twice as large as average worm size. These large programs were removed from the collection to get an equal file size distribution for worms and clean programs.

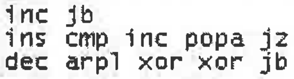
Table 3. File Size Analysis of the Program Collection

Statistic	Worms Size (KB)	Cleans Size (KB)
Average	67	147
Median	33	43
Minimum	1	1
Maximum	762	1968



```
inc     si
jb      short near ptr loc_171+1
ins     word ptr es:[di], dx
cmp     ah, [bx+si]
inc     di
popa
jz      short near ptr loc_16E+1
dec     sp
outsw
arpl    [bp+di+58h], bp
xor     dh, [bx+si]
xor     [bx+si+74h], al
jb      short near ptr loc_178+3
```

Figure 1. Portion of the output of disassembled Netsky.A worm.



```
inc jb
ins cmp inc popa jz
dec arpl xor xor jb
```

Figure 2. Instruction sequences extracted from the disassembled Netsky.A worm.

❖ **Disassembly**

Binaries were transformed to a disassembly representation that is parsed to extract features. The disassembly was obtained using Datarescues' IDA Pro [4]. From these disassembled files we extracted sequences of instructions that served as the primary source for the features in our dataset. A sequence is defined as instructions in succession until a conditional or unconditional branch instruction and/or a function boundary is reached. Instruction sequences thus obtained are of various lengths. We only considered the opcode and the operands were discarded from the analysis. Figure 1 shows a portion of the disassembly of the Netsky.A worm.

❖ **Parsing**

A parser written in PHP translates the disassembly in figure 1 to instruction sequences. Figure 2 displays the output of the parser. Each row in the parsed output represented a single instruction sequence. The raw disassembly of the worm and clean programs resulted in 1972920 instruction sequences. 47% of these sequences belonged to worms while 53% belonged to clean programs.

Feature Extraction

The parsed output was processed through our Feature Extraction Mechanism. Among them 1972920 instruction sequences, 213330 unique sequences were identified with different frequencies of occurrence. We removed the sequences that were found in one class only as they will reduce the classifier to a signature detection technique. This removed 94% of the sequences and only 23738 sequences were found common to both worms and clean programs.

Each sequence was considered as a potential feature.

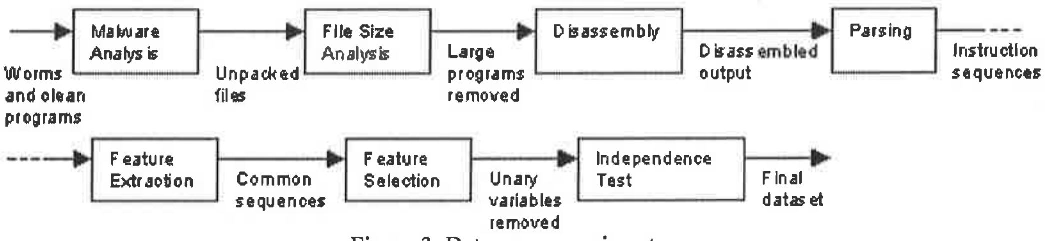


Figure 3. Data preprocessing steps.

VI. Conclusion

Classifier	AUC
C 5.0	99.5
Randam Forest	96.0
Boosting	93.8

In this paper we presented a data mining framework to detect worms. The primary feature used for the process was the frequency of occurrence of variable length instruction sequences. The effect of using such a feature set is twofold as the instruction sequences can be traced back to the original code for further analysis in addition to being used in the classifier. We used the sequences common to both worms and clean programs to remove any biases caused by the features that have all their occurrences in one class only. We showed 99.5% detection rate with a 2.8% false positive rate.

VII. Future Work

The information included for this analysis was extracted from the executable section of the PE file. To achieve a better detection rate this information will be appended from information from other sections of the file. This will include Import Address Table and the PE header. API calls analysis has proven to be an effective tool in malware detection [22]. Moreover header information has been used in heuristic detection [24]. Our next step is to include this information in our feature set.

References

[1] The r project for statistical computing http://www.rproject.org/.

[2] ExEinfo PE. http://www.exeinfo.go.pl/.

[3] Generic Unpacker Win32. http://www.exetools.com/unpackers.htm.

[4] IDA Pro Disassembler. http://www.datarescue.com/idabase/index.htm.

[5] PEiD. http://peid.has.it/.

[6] UPX the Ultimate Packer for eXecutables. http://www.exeinfo.go.pl/.

[7] VMUnpacker. http://dswlab.com/d3.html.

[8] VX Heavens. http://vx.netlux.org.

[9] T. Abou-Assaleh, N. Cercone, V. Keselj, and R. Sweidan. N-gram-based detection of new malicious code. In Proceedings of the 28th Annual International Computer Software and Applications Conference - Workshops and Fast Abstracts - (COMPSAC' 04) - Volume 02, pages 41–42, 2004.

[10] J. Bergeron, M. Debbabi, J. Desharnais, M. M. Erhioui, Y. Lavoie, and N. Tawbi. Static detection of malicious code in executable programs. Symposium on Requirements Engineering for Information Security (SREIS'01), 2001.

[11] L. Breiman. Bagging predictors. Machine Learning, 24(2):123–140, 1996.

[12] L. Breiman. Random forests. Machine Learning, 45(1):5–32, 2001.

[13] F. Cohen. Computer Viruses. PhD thesis, University of Southern California, 1985.

[14] D. Ellis, J. Aiken, K. Attwood, and S. Tenaglia. A behavioral approach to worm detection. In Proceedings of the 2004 ACM Workshop on Rapid Malcode, pages 43–53, 2004.

[15] J. Z. Kolter and M. A. Maloof. Learning to detect malicious executables in the wild. In Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004.

[16] C. Kreibich and J. Crowcroft. Honeycomb creating intrustion detection signatures using honeypots. In 2nd Workshop on Hot Topics in Network, 2003.

[17] R. W. Lo, K. N. Levitt, and R. A. Olsson. Mcf: A malicious code filter. Computers and Security,14(6):541–566, 1995.

[18] J. Nazario. Defense and Detection Strategies against Internet Worms. Van Nostrand Reinhold, 2004.

[19] S. E. Schechter, J. Jung, , and B. A. W. fast detection of scanning worms infections. In Proceedings of Seventh International Symposium on Recent Advances in Intrusion Detection (RAID), 2004.

[20] M. G. Schultz, E. Eskin, E. Zadok, and S. J. Stolfo. Data mining methods for detection of new malicious executables. In Proceedings of the IEEE Symposium on Security and Privacy, pages 38–49, 2001.

[21] M. Siddiqui, M. C. Wang, and J. Lee. Data mining methods for malware detection using instruction sequences. In Proceedings of Artificial Intelligence and Applications, AIA 2008. ACTA Press, 2008.

[22] A. H. Sung, J. Xu, P. Chavez, and S. Mukkamala. Static analyzer of vicious executables. In 20th Annual Computer Security Applications Conference, pages 326–334, 2004.

[23] Symantec. Understanding heuristics: Symantec's bloodhound technology. Technical report, Symantec Corporation, 1997.

[24] P. Szor. The Art of Computer Virus Research and Defense. Addison Wesley for Symantec Press, New Jersey, 2005.

[25] M. Weber, M. Schmid, M. Schatz, and D. Geyer. A toolkit for detecting and analyzing malicious software. In Proceedings of the 18th Annual Computer Security Applications Conference, page 423, 2002.

ISSN (e): 2250 – 3005 || Vol, 04 || Issue, 5 || May – 2014 ||
International Journal of Computational Engineering Research (IJCER)

Efficient Ranking and Suggesting Popular Itemsets In Mobile Stores Using Fp Tree Approach

[1]B.Sujatha Asst prof,[2]Shaista Nousheen Asst.prof, [3]Tasneem rahath Asst prof, [4] Nikhath Fatima Asst.prof

Abstract

We considered the problem of ranking the popularity of items and suggesting popular items based on user feedback. User feedback is obtained by iteratively presenting a set of suggested items, and users selecting items based on their own preferences either the true popularity ranking of items, and suggest true popular items. We consider FP tree approach with some modifications overcoming the complexity that has been seen in other randomized algorithms. The most effective feature of this approach is that it reduces the number of database scans and complexity.

I. INTRODUCTION

1.1 TERMINOLOGY:

In this section we first want to introduce the different terms that we were going to use in our paper as fallows.

1.1.1 Ranking: Ranking is giving rank scores to the most popular item by taking user feedback. The most frequently occurring item is given the highest rank score.

1.1.2 Selection: We focus on the ranking of items where the only available information is the observed selection of items. In learning of the users preference over items, one may leverage some side information about items, but this is out of the scope of this paper.

1.1.3 Imitate: The user study was conducted in very famous mobile stores and which has been used to set of mobiles. The user may check the list and select the set of mobiles which they like most and depending on those like results the new suggestion list has been developed by the algorithm.

1.1.4 Popular: In practice, one may use prior information about item popularity. For example, in the survey the user may select the suggested mobile or they may also select the others. If they selected the already suggested items they will become more popular and if he don't they may get out of the popular list.

1.1.5 Association Rule: Association Rules are if/then statements that help uncover relationships between seemingly unrelated data in the relational database or other information repository. An example of an association rule would be **if a customer buys a nokia mobile, he is 70% interested in also purchasing nokia accessories.**

II. THEORETICAL STUDY

We consider the mobile phone selection and suggesting the best sold mobile and their combinations that were most liked by most of the users. Consider a set of mobiles M: (m1, m2, m3, m4, ….mn) where n > 1. Now we were calculating the set of items in C where were mostly sold and mostly liked by the users, as S

S: (s1, s2, s3, s4, …. sg) where g > 1.

We need to consider an item I, we interpret si as the portion of users that would select item i if suggestions were not made. We assume that the popularity rank scores s as follows:

a) Items of set S were estimated to is as s1 $\geq$ s2 $\geq$ s3 $\geq$ …. sc,

b) s is completely normalized such that it is a probability distribution, i.e., s1 + s2 + s3 + …. +sc = 1. c) si is always positive for all items i.

III. PROPOSED ALGORITHM AND STUDY

We have some of the systems already existing in the same field and we have also identified some of the disadvantages in them as follows:

□ the popularity for any item is given based on the production of that item. This may not give good result because customers may not have the knowledge of true popularity they needed and depend on the results given by the producer.

□ the updates are performed regardless of the true popularity by virtual analysis.

□ Producer have to analyse things manually and complexity involves in this. Due to this time consumption may be high.

□ the algorithms used in this system may fail to achieve true popularity.

We consider the problem learning of the popularity of items that is assumed to be unknown but has to be learned from the observed user's selection of items. We have selected a mobile market and mobile distribution outlets as our data set and examined them completely in all areas where we can give the list of items suggested by the users and we have made web-application to make an survey at real-time and considered the data given by more that 1000 members of different categories of people and applied our proposed Fp tree approach on the set of data and started suggesting the item in the mobile outlets for the actual users, which had helped the mobile phone companies and also the outlet in-charges. We have implemented the same in some of the mobile outlets in INDIA where we got very good results. The actual goal of the system is to efficiently learn the popularity of items and suggest the popular items to users. This was done to the user to suggest them the mostly used mobiles and their accessories, such that they also buy the best and at the same time the outlet owner will also get benefited. The most important feature in our project is suggesting the users by refreshing the latest results every time the user gives the input and changes his like list.

Now we have overcome many of the disadvantages of the existing systems and achieved many advantages with the proposed algorithm and method as follows:

□ In our approach, we consider the problem of ranking the popularity of items and suggesting popular items based on user feedback.

□ User feedback is obtained by iteratively presenting a set of suggested items, and users selecting items based on their own preferences either from this suggestion set or from the set of all possible items.

□ The goal is to quickly learn the true popularity ranking of items and suggest true popular items.

□ In this system better algorithms are used. The algorithms use ranking rules and suggestion rules in order to achieve true popularity.

IV. PROPOSED APPROACHFP-TREE

Like most traditional studies in association mining, we de_ne the frequent pattern mining problem as follows.De_nition 1 (Frequent pattern) Let I = fa1; a2; : : :; amg be a set of items, and a transaction database DB = hT1; T2; : : :; Tni, where Ti (i 2 [1::n]) is a transaction which contains a set of items in I. The support1(or occurrence frequency) of a pattern A, which is a set of items, is the number of transactions containing A in DB. A, is a frequent pattern if A's support is no less than a prede_ned minimum support threshold, _. 2Given a transaction database DB and a minimum support threshold, _, the problem of _nding the complete set of frequent patterns is called the frequent pattern mining problem.

2.1 Frequent Pattern Tree

To design a compact data structure for efficient frequent pattern mining, let's _rst examine a tiny example.Example 1 Let the transaction database, DB, be (the _rst two columns of) Table 1 and the minimum supportthreshold be 3.

TID	Items bought	(Ordered)frequent items
100	f, a, c, d, g, i,m,p	f, c, a,m, p
200	a, b, c, f, l,m,o	f, c, a, b,m
300	b, f, h, j,o	f, b
400	b, c, k, s,p	c, b, p
500	a, f, c, e, l, p,m,n	f, c, a,m, p

A compact data structure can be designed based on the following observations:

[1]. Since only the frequent items will play a role in the frequent-pattern mining, it is necessary to perform one scan of transaction database *DB* to identify the set of frequent items (with *frequency count* obtained as a by-product).

[2]. If the *set* of frequent items of each transaction can be stored in some compact structure,it may be possible to avoid repeatedly scanning the original transaction database.

[3]. If multiple transactions share a set of frequent items, it may be possible to merge the shared sets with the number of occurrences registered as *count*. It is easy to check whether two sets are identical if the frequent items in all of the transactions are listed according to a fixed order.

[4]. If two transactions share a common prefix, according to some sorted order of frequent items, the shared parts can be merged using one prefix structure as long as the *count* is registered properly. If the frequent items are sorted in their *frequency descending order*,there are better chances that more prefix strings can be shared.With the above observations, one may construct a frequent-pattern tree as follows.

First, a scan of *DB* derives a *list* of frequent items, _(*f* :4), (*c*:4), (*a*:3), (*b*:3), (*m*:3), (*p*:3)_(the number after ":" indicates the support), in which items are ordered in frequency descending rder. This ordering is important since each path of a tree will follow this order. For convenience of later discussions, the frequent items in each transaction are listed in this ordering in the rightmost column of Table 1.Second, the root of a tree is created and labeled with "*null*". The FP-tree is constructed as follows by scanning the transaction database *DB* the second time.

[1]. The scan of the first transaction leads to the construction of the first branch of the tree:

[2]. _(*f* :1), (*c*:1), (*a*:1), (*m*:1), (*p*:1)_. Notice that the frequent items in the transaction are listed according to the order in the *list* of frequent items.

[3]. For the second transaction, since its (ordered) frequent item list _*f, c, a, b,m*_ shares a

[4]. common prefix _*f, c, a*_ with the existing path _*f, c, a,m, p*_, the count of each node along the prefix is incremented by 1, and one new node (*b*:1) is created and linked as a child of (*a*:2) and another new node (*m*:1) is created and linked as the child of (*b*:1).

[5]. For the third transaction, since its frequent item list _*f, b*_ shares only the node _*f*_ with

[6]. the *f* -prefix subtree, *f* 's count is incremented by 1, and a new node (*b*:1) is created and linked as a child of (*f* :3).

[7]. The scan of the fourth transaction leads to the construction of the second branch of the tree, _(*c*:1), (*b*:1), (*p*:1)_.

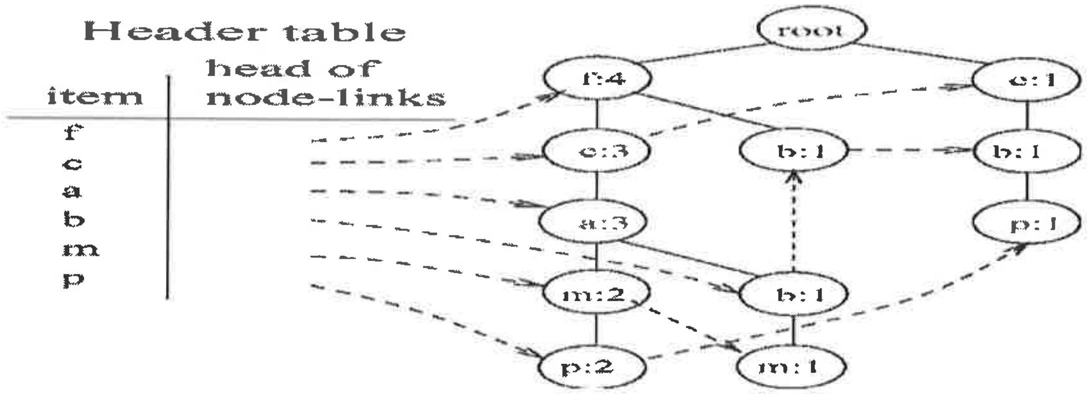
[8]. For the last transaction, since its frequent item list _*f, c, a,m, p*_ is identical to the first one, the path is shared with the count of each node along the path incremented by 1.

[9]. To facilitate tree traversal, an item header table is built in which each item points to its first occurrence in the tree via a node-link. Nodes with the same item-name are linked in sequence via such *node-links*. After scanning all the transactions, the tree, together with the associated node-links, are shown in figure 1.

Based on this example, a *frequent-pattern tree* can be designed as follows. *Definition 1* (*FP-tree*). A *frequent-pattern tree* (or *FP-tree* in short) is a tree structure defined below.

[1]. It consists of one root labeled as "*null*", a set of item-prefix sub trees as the children of the root, and a frequent-item-header table.

[2]. Each node in the item-prefix sub tree consists of three fields: *item-name*, *count*, and *node-link*, where *item-name* registers which item this node represents, *count* registers the number of transactions represented by the portion of the path reaching this node, and



node-link links to the next node in the FP-tree carrying the same item-name, or null if there is none.

[3] Each entry in the frequent-item-header table consists of two fields, (1) *item-name* and (2) *head of node-link* (a pointer pointing to the first node in the FP-tree carrying the *item-name*).Based on this definition, we have the following FP-tree construction algorithm.

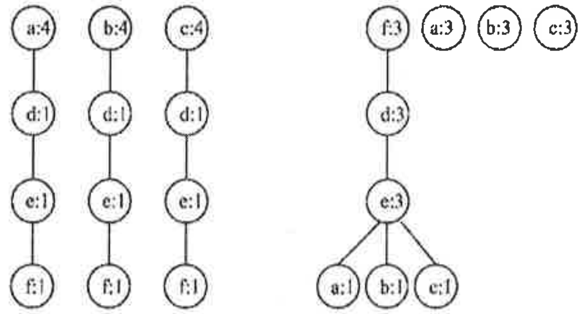
Algorithm 1 (FP-tree construction).
Input: A transaction database *DB* and a minimum support threshold ξ.
Output: FP-tree, the frequent-pattern tree of *DB*.
Method: The FP-tree is constructed as follows.
1. Scan the transaction database *DB* once. Collect *F*, the set of frequent items, and the support of each frequent item. Sort *F* in support-descending order as *FList*, the *list* of frequent items.
2. Create the root of an FP-tree, *T* , and label it as "null". For each transaction *Trans* in *DB* do the following. Select the frequent items in *Trans* and sort them according to the order of *FList*. Let the sorted frequent-item list in *Trans* be [*p* | *P*], where *p* is the first element and *P* is theremaining list. Call *insert tree*([*p* | *P*], *T*). The function *insert tree*([*p* | *P*], *T*) is performed as follows. If *T* has a child *N* such that *N.item-name* = *p.item-name*, then increment *N*'s count by 1; else create a new node *N*, with its count initialized to 1, its parent link linked to *T* , and its node-link linked to the nodes with the same *item-name* via the node-link structure. If *P* is nonempty, call *nserttree(P, N)* recursively. Analysis. The FP-tree construction takes exactly two scans of the transaction database: The first scan collects the set of frequent items, and the second scan constructs the FP-tree. The cost of inserting a transaction *Trans* into the FP-tree is (|*freq*(*Trans*)|),where*freq*(*Trans*) is the set of frequent items in *Trans*. We will show that the FP-tree contains the complete information for frequent-pattern mining.

Completeness and compactness of FP-tree

There are several important properties of FP-tree that can be derived from the FP-tree construction process. Given a transaction database *DB* and a support threshold ξ . Let *F* be the frequent items in *DB*. For each transaction *T* , *freq*(*T*) is the set of frequent items in *T* , i.e., *freq*(*T*) = *T* ∩ *F*,and is called the *frequent item projection* of transaction *T* . According to the *Apriori* principle, the set of frequent item projections of transactions in the database is sufficient for mining the complete set of frequent patterns, because an infrequent item plays no role in frequent patterns.

Lemma 1. *Given a transaction database DB and a support threshold ξ, the complete set of frequent item projections of transactions in the database can be derived from DB's FP-tree.*Rationale. Based on the FP-tree construction process, for each transaction in the *DB*, its frequent item projection is mapped to one path in the FP-tree.For a path *a*1*a*2 . . . *ak* from the root to a node in the FP-tree, let *cak* be the count at the node labeled *ak* and *c_ak* be the sum of counts of children nodes of *ak* . Then, according to the construction of the FP-tree, the path registers frequent item projections of *cak*– *c_ak* transactions.Therefore, the FP-tree registers the complete set of frequent item projections without duplication.Based on this lemma, after an FP-tree for *DB* is constructed, it contains the complete information for mining frequent patterns from the transaction database. Thereafter, only the FP-tree is needed in the remaining mining process, regardless of the number and length of the frequent patterns.

Lemma 2. *Given a transaction database DB and a support threshold ξ. Without consideringthe (null) root, the size of an FP-tree is bounded by _T∈DB |freq(T)|, and the height of the tree is bounded by* max*T∈DB{|freq(T)|}, where freq(T) is the frequent item projection of transaction T* Rationale. Based on the FP-tree construction process, for any transaction *T* in *DB*, there exists a path in the FP-tree starting from the corresponding item prefix subtree so that the set of nodes in the path is exactly the same set of frequent items in *T* . The root is the only extra node that is not created by frequent-item insertion, and each node contains one node-link and one count. Thus we have the bound of the size of the tree stated in the Lemma.The height of any *p*-prefix subtree is the maximum number of frequent items in any transaction with *p* appearing at the head of its frequent item list. Therefore, the height of the tree is bounded by the maximal number of frequent items in any transaction in the database, if we do not consider the additional level added by the root.Lemma 2.2 shows an important benefit of FP-tree: the size of an FP-tree is bounded by the size of its corresponding database because each transaction will contribute at most one path to the FP-tree, with the length equal to the number of frequent items in that transaction. Since there are often a lot of sharings of frequent items among transactions, the size of the tree is usually much smaller than its original database. Unlike the *Apriori*-like method which may generate an exponential number of candidates in the worst case, under no circumstances, may an FP-tree with an exponential number of nodes be generated.FP-tree is a highly compact structure which stores the information for frequent-patternmining. Since a single path "*a*1 → *a*2 → · · · → *an*" in the *a*1-prefix subtree registers all the transactions whose maximal frequent set is in the form of "*a*1 → *a*2 → · · →*ak*" for any 1 ≤ *k* ≤ *n*, the size of the FP-tree is substantially smaller than the size of the database and that of the candidate sets generated in the association rule mining.The items in the frequent item set are ordered in the support-descending order: More frequently occurring items are more likely to be shared and thus they are arranged closer to the top of the FP-tree. This ordering enhances the compactness of the FP-tree structure. However, this does not mean that the tree so constructed *always* achieves the maximal compactness. With the knowledge of particular data characteristics, it is sometimes possible to achieve even better compression than the frequency-descending ordering. Consider the following example. Let the set of transactions be: {*adef* , *bdef* , *cdef* , *a*, *a*, *a*, *b*, *b*, *b*, *c*, *c*, *c*}, and the minimum support threshold be 3. The frequent item set associated with support count becomes {*a*:4, *b*:4, *c*:4, *d*:3, *e*:3, *f* :3}. Following the item frequency ordering *a* → *b* → *c* → *d* → *e* → *f* , the FP-tree constructed will contain 12 nodes, as shown in figure 2(a). However, following another item ordering *f* → *d* → *e* → *a* → *b* → *c*, it will contain only 9 nodes, as shown in figure 2(b).The compactness of FP-tree is also verified by our experiments. Sometimes a rather small FP-tree is resulted from a quite large database. For example, for the database *Connect-4* used in *MaxMiner* (Bayardo, 1998), which contains 67,557 transactions with 43 items in each transaction, when the support threshold is 50% (which is used in the *MaxMiner* experiments (Bayardo, 1998)), the total number of occurrences of frequent items is 2,219,609, whereas the total number of nodes in the FP-tree is 13,449 which represents a reduction ratio of 165.04, while it still holds hundreds of thousands of frequent patterns! (Notice that for databases with mostly short transactions, the reduction ratio is not that high.)



a) FPtree follows the support ordering b) FPtree does not follow the support ordering

Therefore,it is not surprising some gigabyte transaction database containing many long patterns may even generate an FP-tree that fits in main memory. Nevertheless, one cannot assume that an FP-tree can always fit in main memory no matter how large a database is. Methods for highly scalable *FP-growth* mining will be discussed in Section 5.

4.1 Mining frequent_patterns using FP-tree: Construction of a compact FP-tree ensures that subsequent mining can be performed with a rather compact data structure. However, this does not automatically guarantee that it will be highly efficient since one may still encounter the combinatorial problem of candidate generation if one simply uses this FP-tree to generate and check all the candidate patterns. In this section, we study how to explore the compact information stored in an FP-tree, develop the principles of frequent-pattern growth by examination of our running example, explore how to perform further optimization when there exists a single prefix path in an FP-tree, and propose a frequent-pattern growth algorithm, *FP-growth,* for mining the *complete set of frequent patterns* using FP-tree. *4.1.1 Principles of frequent-pattern growth for FP-tree mining* In this subsection, we examine some interesting properties of the FP-tree structure which will facilitate frequent-pattern mining.

Property 1 (*Node-link property*). *For any frequent item ai , all the possible patterns containing only frequent items and ai can be obtained by following ai 's node-links, starting from ai 's head in the FP-tree header.*This property is directly from the FP-tree construction process, and it facilitates the access of all the frequent-pattern information related to *ai* by traversing the FP-tree once following *ai* 's node-links.To facilitate the understanding of other properties of FP-tree related to mining, we first go through an example which performs mining on the constructed FP-tree (figure 1) in Example 1. *Example 2.* Let us examine the mining process based on the constructed FP-tree shown in figure 1. Based on Property 3.1, all the patterns containing frequent items that a node *ai*participates can be collected by starting at *ai* 's node-link head and following its node-links.
We examine the mining process by starting from the bottom of the node-link header table. For node *p*, its immediate frequent pattern is (*p*:3), and it has two paths in the FP-tree:_*f* :4, *c*:3, *a*:3,*m*:2, *p*:2_ and _*c*:1, *b*:1, *p*:1_. The first path indicates that string"(*f, c, a,m, p*)" appears twice in the database. Notice the path also indicates that string _*f, c, a*_ appears three times and _*f*_ itself appears even four times. However, they only appear twice *together* with *p*. Thus, to study which string appear together with *p*, only *p*'s prefix path _*f* :2, *c*:2, *a*:2,*m*:2_ (or simply, _*f cam*:2_) counts. Similarly, the second path indicates string "(*c, b, p*)" appears once in the set of transactions in *DB*, or *p*'s prefix pathis _*cb*:1_. These two prefix paths of *p*, "*{*(*f cam*:2), (*cb*:1)*}*", form *p*'s subpattern-base, which is called *p*'s conditional pattern base (i.e., the subpattern-base under the condition of *p*'s existence). Construction of an FP-tree on this conditional pattern-base (which is called *p*'s conditional FP-tree) leads to only one branch (*c*:3). Hence, only one frequent pattern (*cp*:3) is derived. (Notice that a pattern is an itemset and is denoted by a string here.) The search for frequent patterns associated with *p* terminates. For node *m*, its immediate frequent pattern is (*m*:3), and it has two paths, _*f* :4, *c*:3, *a*:3,*m*:2_ and _*f* :4, *c*:3, *a*:3, *b*:1, *m*:1_. Notice *p* appears together with *m* as well, however, there is no need to include *p* here in the analysis since any frequent patterns involving *p* has been analyzed in the previous examination of *p*. Similar to the above analysis, *m*'s conditional pattern-base is *{*(*fca*:2), (*fcab*:1)*}*. Constructing an FP-tree on it, we derive *m*'s conditional FP-tree, _ *f* :3, *c*:3, *a*:3_, a single frequent pattern path, as shown in figure 3. This conditional FP-tree is then mined recursively by calling *mine*(_*f* :3, *c*:3, *a*:3_ |*m*). Figure 3 shows that "*mine*(_*f* :3, *c*:3, *a*:3_ |*m*)" involves mining three items (*a*), (*c*), (*f*) in sequence. The first derives a frequent pattern (*am*:3), a conditional pattern-base *{*(*fc*:3)*}*, and then a call "*mine*(_ *f* :3, *c*:3_ | *am*)"; the second derives a frequent pattern (*cm*:3), a conditional pattern-base *{*(*f* :3)*}*, and then a call "*mine*(_ *f* :3_ | *cm*)"; and the third derives only a frequent pattern (*fm*:3). Further recursive call of "*mine*(_ *f* :3, *c*:3_ | *am*)" derives two patterns (*cam*:3) and (*fam*:3), and a conditional pattern-base *{*(*f* :3)*}*, which then leads to a call "*mine*(_ *f* :3_ | *cam*)", that derives the longest pattern (*fcam*:3). Similarly, the call of "*mine*(_ *f* :3_ | *cm*)" derives one pattern (*fcm*:3). Therefore, the set of frequent patterns involving *m* is *{*(*m*:3), (*am*:3), (*cm*:3), (*f m*:3), (*cam*:3), (*fam*:3), (*fcam*:3), (*fcm*:3)*}*. This indicates that *a single path FP-tree can be mined by outputting all the combinations of the items in the path.*Similarly, node *b* derives (*b*:3) and it has three paths: _ *f* :4, *c*:3, *a*:3, *b*:1_, _ *f* :4, *b*:1_, and _*c*:1, *b*:1_. Since *b*'s conditional pattern-base *{*(*fca*:1), (*f* :1), (*c*:1)*}* generates no frequent item, the mining for *b* terminates. Node *a* derives one frequent pattern *{*(*a*:3)*}* and one subpattern base *{*(*f c*:3)*}*, a single-path conditional FP-tree. Thus, its set of frequent pattern

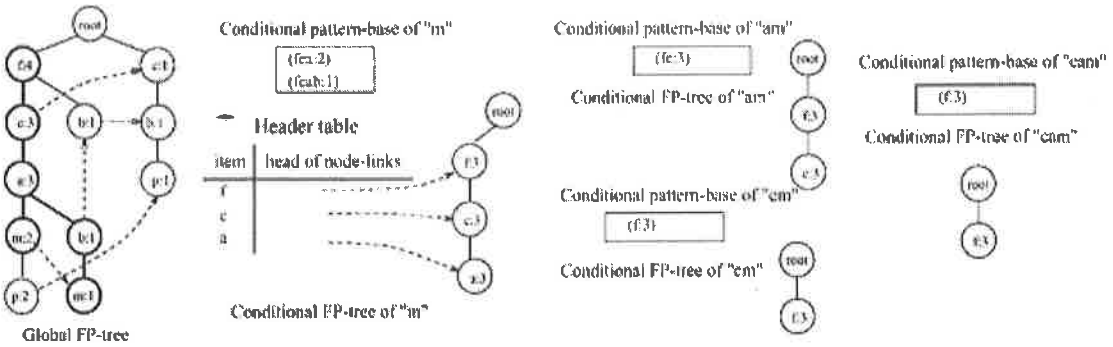

root
f:4
c:1
c:3
b:1
b:1
a:3
p:1
m:2
b:1
p:2
m:1
Global FP-tree
Conditional pattern-base of "m"
{fca:2}
{fcab:1}
Header table
item
head of node-links
f
c
a
root
f:3
c:3
a:3
Conditional FP-tree of "m"
Conditional pattern-base of "am"
(fc:3)
root
Conditional FP-tree of "am"
f:3
c:3
Conditional pattern-base of "cm"
(f:3)
Conditional FP-tree of "cm"
root
f:3
Conditional pattern-base of "cam"
(f:3)
Conditional FP-tree of "cam"
root
f:3

Table 2. Mining frequent patterns by creating conditional (sub)pattern-bases.

Item	Conditional pattern-base	Conditional FP-tree
P	{(*f cam*:2), (*cb*:1)}	{(*c*:3)}|*p*
M	{(*f ca*:2), (*fcab*:1)}	{(*f*:3, *c*:3, *a*:3)}|*m*
b	*{*(*f ca*:1), (*f* :1), (*c*:1)*}*	∅
a	*{*(*f c*:3)*}*	{(*f*:3, *c*:3)}|*a*
c	*{*(*f* :3)*}*	{(*f*:3)}|*c*
f	∅	∅

can be generated by taking their combinations. Concatenating them with (*a*:3), we have *{*(*f a*:3), (*ca*:3), (*fca*:3)*}*. Node *c* derives (*c*:4) and one subpattern-base *{*(*f* :3)*}*, and the set of frequent patterns associated with (*c*:3) is *{*(*fc*:3)*}*. Node *f* derives only (*f* :4) but no conditional pattern-base.The conditional pattern-bases and the conditional FP-trees generated are summarized in Table 2.The correctness and completeness of the process in Example 2 should be justified.This is accomplished by first introducing a few important properties related to the mining process.

Property 2 (*Prefix path property*). *To calculate the frequent patterns with suffix ai , onlythe* prefix subpathes *of nodes labeled ai in the FP-tree need to be accumulated, and the frequency count of every node in the prefix path should carry the same count as that in the corresponding node ai in the path.*Rationale. Let the nodes along the path *P* be labeled as *a*1, . . . , *an* in such an order that *a*1 is the root of the prefix subtree, *an* is the leaf of the subtree in *P*, and *ai* (1 ≤ *i* ≤ *n*) is the node being referenced. Based on the process of FP-tree construction presented in Algorithm 1, for each prefix node *ak* (1 ≤ *k* < *i*), the prefix subpath of the node *ai* in *P* occurs together with *ak* exactly *ai .count* times. Thus every such prefix node should carry the same count as node *ai* . Notice that a postfix node *am* (for *i* < *m* ≤ *n*) along the same path also co-occurs with node *ai*. However, the patterns with *am* will be generated when examining the suffix node *am*, enclosing them here will lead to redundant generation of the patterns that would have been generated for *am*. Therefore, we only need to examine the prefix subpath of *ai* in *P*. For example, in Example 2, node *m* is involved in a path _*f* :4, *c*:3, *a*:3,*m*:2, *p*:2_, to calculate the frequent patterns for node *m* in this path, only the prefix subpath of node *m*, which is _*f* :4, *c*:3, *a*:3_, need to be extracted, and the frequency count of every node in the prefix path should carry the same count as node *m*. That is, the node counts in the prefix path should be adjusted to _*f* :2, *c*:2, *a*:2_. Based on this property, the prefix subpath of node *ai* in a path *P* can be copied and transformed into a count-adjusted prefix subpath by adjusting the frequency count of every node in the prefix subpath to the same as the count of node *ai* . The prefix path so transformed is called the *transformed prefix path* of *ai* for path *P*. Notice that the set of transformed prefix paths of *ai* forms a small database of patterns which co-occur with *ai* . Such a database of patterns occurring with *ai* is called *ai 's conditional pattern-base*, and is denoted as "*pattern base* | *ai* ". Then one can compute all the frequent patterns associated with *ai* in this *ai* -conditional pattern-base by creating a small FP-tree, called *ai 's conditional FP-tree* and denoted as "FP-tree | *ai* ". Subsequent mining can be performed on this small conditional FP-tree. The processes of construction of conditional pattern-bases and conditional FP-trees have been demonstrated in Example 2. This process is performed recursively, and the frequent patterns can be obtained by a pattern-growth method, based on the following lemmas and corollary.

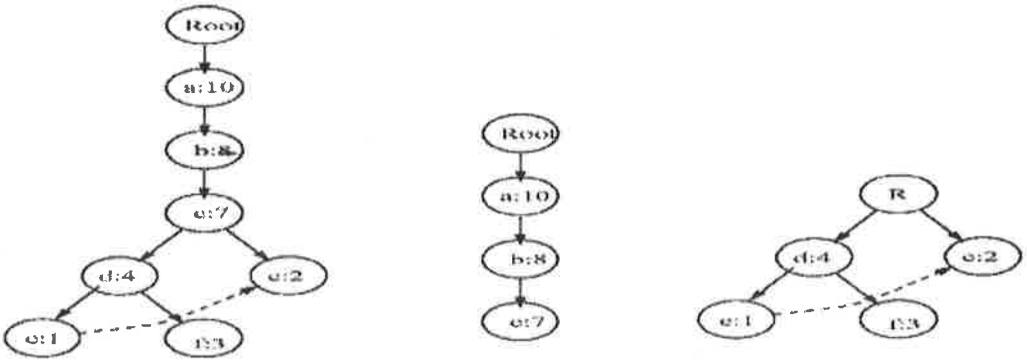
Lemma 1 (*Fragment growth*). *Let α be an itemset in DB, B be α's conditional patternbase, and β be an itemset in B. Then the support of α* ∪*β in DB is equivalent to the support of β in B.* Rationale. According to the definition of conditional pattern-base, each (sub)transaction in *B* occurs under the condition of the occurrence of *α* in the original transaction database *DB*. If an itemset *β* appears in *B ψ* times, it appears with *α* in *DBψ* times as well. Moreover, since all such items are collected in the conditional pattern-base of *α*, *α* ∪ *β* occurs exactly *ψ* times in *DB* as well. Thus we have the lemma. From this lemma, we can directly derive an important corollary.

Corollary 1 (*Pattern growth*). *Let α be a frequent itemset in DB, B be α's conditional pattern-base, and β be an itemset in B. Then α* ∪ *β is frequent in DB if and only if β is frequent in B.* Based on Corollary 3.1, mining can be performed by first identifying the set of frequent 1-itemsets in *DB*, and then for each such frequent 1-itemset, constructing its conditional pattern-bases, and mining its set of frequent 1-itemsets in the conditional pattern-base, and so on. This indicates that the process of mining frequent patterns can be viewed as first mining frequent 1-itemset and then progressively growing each such itemset by mining its conditional pattern-base, which can in turn be done similarly. By doing so, a frequent *k*-itemset mining problem is successfully transformed into a sequence of *k* frequent 1- itemset mining problems via a set of conditional pattern-bases. Since mining is done by pattern growth, there is no need to generate any candidate sets in the entire mining process. Notice also in the construction of a new FP-tree from a *conditional pattern-base* obtained during the mining of an FP-tree, the items in the frequent itemset should be ordered in the frequency descending order of *node occurrence* of each item instead of its *support* (which represents item occurrence). This is because each node in an FP-tree may represent many occurrences of an item but such a node represents a single unit (i.e., the itemset whose elements always occur together) in the construction of an item-associated FP-tree.

3.2. Frequent-pattern growth with single prefix path of FP-tree

The frequent-pattern growth method described above works for all kinds of FP-trees. However, further optimization can be explored on a special kind of FP-tree, called *single prefixpathFP-tree*, and such an optimization is especially useful at mining long frequent patterns. A single prefix-path FP-tree is an FP-tree that consists of only a single path or a single prefix path stretching from the root to the first branching node of the tree, where a *branching node* is a node containing more than one child. Let us examine an example.

Example 3. Figure 4(a) is a single prefix-path FP-tree that consists of one prefix path,_(*a*:10)→(*b*:8)→(*c*:7)_, stretching from the root of the tree to the first branching node (*c*:7).Although it can be mined using the frequent-pattern growth method described above, a bettermethod is to split the tree into two fragments: the single prefix-path, _(*a*:10)→(*b*:8)→(*c*:7)_, as shown in figure 4(b), and the multipath part, with the root replaced by a pseudoroot *R*, as shown in figure 4(c). These two parts can be mined separately and then combined together.Let us examine the two separate mining processes. All the frequent patterns associated with the first part, the single prefix-path *P* = _(*a*:10)→(*b*:8)→(*c*:7)_, can be mined by enumeration of all the combinations of the subpaths of *P* with the support set to the minimum support of the items contained in the subpath. This is because each such subpath is distinct and occurs the same number of times as the *minimum occurrence frequency among the items in the subpath* which is equal to the support of the last item in the subpath. Thus, path *P* generates the following set of frequent patterns, *freq pattern set(P)* = {(*a*:10), (*b*:8), (*c*:7), (*ab*:8), (*ac*:7), (*bc*:7), (*abc*:7)}.Let *Q* be the second FP-tree (figure 4(c)), the multipath part rooted with *R*. *Q* can be mined as follows.First, *R* is treated as a *null* root, and *Q* forms a multipath FP-tree, which can be mined using a typical frequent-pattern growth method. The mining result is: *freq pattern set*(*Q*)= {(*d*:4), (*e*:3), (*f* :3), (*d f* :3)}. *Figure*



(a) Single prefix-path tree (b) Single-path portion P (c) Multipath portion Q

Figure 4. Mining an FP-tree with a single prefix path.

Second, for each frequent itemset in *Q*, *R* can be viewed as a conditional frequent pattern-base, and each itemset in *Q* with each pattern generated from *R* may form a distinct frequent pattern. For example, for (*d*:4) in *freq pattern set*(*Q*), *P* can be viewed as its conditional pattern-base, and a pattern generated from *P*, such as (*a*:10), will generate with it a new frequent itemset, (*ad*:4), since *a* appears together with *d* at most four times. Thus, for (*d*:4) the set of frequent patterns generated will be (*d*:4)×*freq pattern set*(*P*) = {(*ad*:4), (*bd*:4), (*cd*:4), (*abd*:4), (*acd*:4), (*bcd*:4), (*abcd*:4)}, where *X* × *Y* means that every pattern in *X* is combined with everyone in *Y* to form a "cross-product-like" larger itemset with the support being the minimum support between the two patterns. Thus, the complete set of frequent patterns generated by combining the results of *P* and *Q* will be *freq pattern set*(*Q*)×*freq pattern set*(*P*), with the support being the support of the itemset in *Q* (which is always no more than the support of the itemset from *P*). In summary, the set of frequent patterns generated from such a single prefix path consists of three distinct sets: (1) *freq pattern set*(*P*), the set of frequent patterns generated from the single prefix-path, *P*; (2) *freq pattern set*(*Q*), the set of frequent patterns generated from the multipath part of the FP-tree, *Q*; and (3) *freq pattern set*(*Q*)×*freq pattern set*(*P*), the set of frequent patterns involving both parts. We first showif an FP-tree consists of a single path *P*, one can generate the set of frequent patterns according to the following lemma.

Lemma 2 (*Pattern generation for an FP-tree consisting of single path*). *Suppose an FP-tree T consists of a single path P. The complete set of the frequent patterns of T can be generated by enumeration of all the combinations of the subpaths of P with the support being the minimum support of the items contained in the subpath.* Rationale. Let the single path *P* of the FP-tree be _*a*1:*s*1→*a*2:*s*2→ ·· ·→*ak* :*sk* _. Since the FP-tree contains a single path *P*, the support frequency *si* of each item *ai* (for 1 ≤ *i* ≤ *k*) is the frequency of *ai* co-occurring with its prefix string. Thus, any combination of the items in the path, such as _*ai* , . . . , *aj* _ (for 1 ≤ *i*, *j* ≤ *k*), is a frequent pattern, with their cooccurrence frequency being the minimum support among those items. Since every item in each path *P* is unique, there is no redundant pattern to be generated with such a combinational generation. Moreover, no frequent patterns can be generated outside the FP-tree. Therefore, we have the lemma. We then show if an FP-tree consists of a single prefix-path, the set of frequent patterns can be generated by splitting the tree into two according to the following lemma.

Lemma 3 (*Pattern generation for an FP-tree consisting of single prefix path*). *Suppose an FP-tree T, similar to the tree in figure* 4(a), *consists of* (1) *a single prefix path P, similar to the tree P in figure* 4(b), *and* (2) *the multipath part, Q, which can be viewed as an independent FP-tree with a pseudo-root R, similar to the tree Q in figure* 4(c). *The complete set of the frequent patterns of T consists of the following three portions:*

1. *The set of frequent patterns generated from P by enumeration of all the combinations of the items along path P, with the support being the minimum support among all the items that the pattern contains.*
2. *The set of frequent patterns generated from Q by taking root R as "null."*
3. *The set of frequent patterns combining P and Q formed by taken the cross-product of the frequent patterns enerated from P and Q, denoted as freq pattern set*(*P*) × *freq pattern set*(*Q*), *that is, each frequent itemset is the union of one frequent itemset from P and one from Q and its support is the minimum one between the supports of the two itemsets.* Rationale. Based on the FP-tree construction rules, each node *ai* in the single prefix path of the FP-tree appears only once in the tree. The single prefix-path of the FP-tree forms a new FP-tree *P*, and the multipath part forms another FP-tree *Q*. They do not share nodes representing the same item. Thus, the two FP-trees can be mined separately. First, we show that each pattern generated from one of the three portions by llowing the pattern generation rules is distinct and frequent. According to Lemma 3.2, each pattern generated from *P*, the FP-tree formed by the single prefix-path, is distinct and frequent. The set of frequent patterns generated from *Q* by taking root *R* as "null" is also distinct and frequent since such patterns exist without combining any items in their conditional databases (which are in the items in *P*. The set of frequent patterns generated by combining *P* and *Q*, that is, taking the cross-product of the frequent patterns generated from *P* and *Q*, with the support being the minimum one between the supports of the two itemsets, is also distinct and frequent. This is because each frequent pattern generated by *P* can be considered as a frequent pattern in the conditional pattern-base of a frequent item in *Q*, and whose support should be the minimum one between the two supports since this is the frequency that both patterns appear together.Second, we show that no patterns can be generated out of this three portions. Sinceaccording to Lemma 3.1, the FP-tree *T* without being split into two FP-trees *P* and *Q* generatesthe complete set of frequent patterns by pattern growth. Since each pattern generated from *T* will be generated from either the portion *P* or *Q* or their combination, the method generates the complete set of frequent patterns. *The frequent-pattern growth algorithm* Based on the above lemmas and properties, we have the following algorithm for mining frequent patterns using FP-tree.

Algorithm 2 (FP-growth: *Mining frequent patterns with FP-tree by pattern fragment growth*).

Input: A database *DB*, represented by FP-tree constructed according to Algorithm 1, and a minimum support threshold ξ.

Output: The complete set of frequent patterns.68 HAN ET AL.
Method: *call FP-growth*(FP-tree, *null*).

```
Procedure FP-growth(Tree, $\alpha$)
{
(1) if Tree contains a single prefix path // Mining single prefix-path FP-tree
(2) then {
(3) let P be the single prefix-path part of Tree;
(4) let Q be the multipath part with the top branching node replaced by a null root;
(5)for each combination (denoted as $\beta$) of the nodes in the path P do
(6) generate pattern $\beta \cup \alpha$ with support = minimum support of nodes in $\beta$;
(7) let freq pattern set(P) be the set of patterns so generated; }
(8) else let Q be Tree;
(9)for each item ai in Q do { // Mining multipath FP-tree
(10) generate pattern $\beta = ai \cup \alpha$ with support = ai .support;
(11) construct $\beta$'s conditional pattern-base and then $\beta$'s conditional FP-tree Tree$\beta$ ;
(12) if Tree$\beta = \emptyset$
(13) then call FP-growth(Tree$\beta$, $\beta$);
(14) let freq pattern set(Q) be the set of patterns so generated; }
(15) return(freq pattern set(P) $\cup$ freq pattern set(Q) $\cup$ (freq pattern set(P)
$\times$freq pattern set(Q)))
}
```

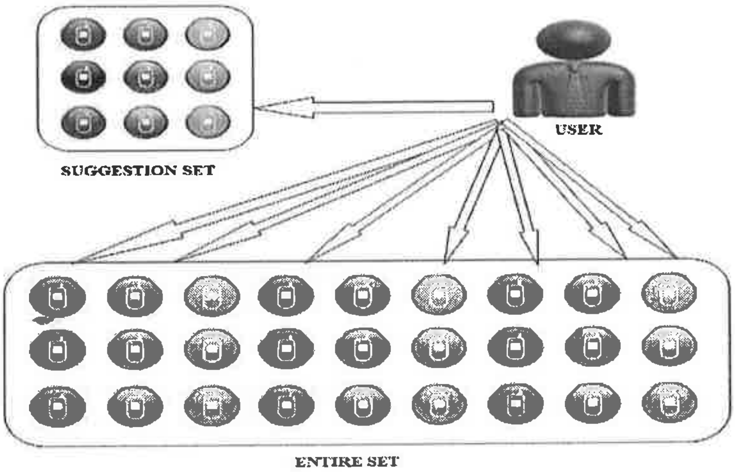
Analysis. With the properties and lemmas in Sections 2 and 3, we show that the algorithm correctly finds the complete set of frequent itemsets in transaction database *DB*. As shown in Lemma 2.1, FP-tree of *DB* contains the complete information of *DB* in relevance to frequent pattern mining under the support threshold ξ . If an FP-tree contains a single prefix-path, according to Lemma 3.3, the generation of the complete set of frequent patterns can be partitioned into three portions: the single prefix-path portion *P*, the multipath portion *Q*, and their combinations. Hence we have lines (1)-(4) and line (15) of the procedure. According to Lemma 3.2, the generated patterns for the single prefix path are the enumerations of the subpaths of the prefix path, with the support being the minimum support of the nodes in the subpath. Thus we have lines (5)-(7) of the procedure.

After that, one can treat the multipath portion or the FP-tree that does not contain the single prefix-path as portion *Q* (lines (4) and (8)) and construct conditional pattern-base and mine its conditional FP-tree for each frequent itemset *ai* . The correctness and completeness of the prefix path transformation are shown in Property 3.2. Thus the conditional pattern-bases store the complete information for frequent pattern mining for *Q*. According to Lemmas 3.1 and its corollary, the patterns successively grown from the conditional FP-trees are the set of sound and complete frequent patterns. Especially, according to the fragment growth property, the support of the combined fragments takes the support of the frequent itemsets generated in the conditional pattern-base. Therefore, we have lines (9)-(14) of the procedure. Line (15) sums up the complete result according to Lemma 3.3. Let's now examine the efficiency of the algorithm. The *FP-growth* mining process scans the FP-tree of *DB* once and generates a small pattern-base *Bai* for each frequent item *ai* , each consisting of the set of transformed prefix paths of *ai* . Frequent pattern mining is then recursively performed on the small pattern-base *Bai* by constructing a conditional FP-tree for *Bai*. As reasoned in the analysis of Algorithm 1, an FP-tree is usually much smaller than the size of *DB*. Similarly, since the conditional FP-tree, "FP-tree | *ai* ", is constructed on the pattern-base *Bai* , it should be usually much smaller and never bigger than *Bai* . Moreover, a pattern-base *Bai* is usually much smaller than its original FP-tree, because it consists of the transformed prefix paths related to only one of the frequent items, *ai* . Thus, each subsequent mining process works on a set of usually much smaller pattern-bases and conditional FPtrees. Moreover, the mining operations consist of mainly prefix count adjustment, counting local frequent items, and pattern fragment concatenation. This is much less costly than generation and test of a very large number of candidate patterns. Thus the algorithm is efficient.

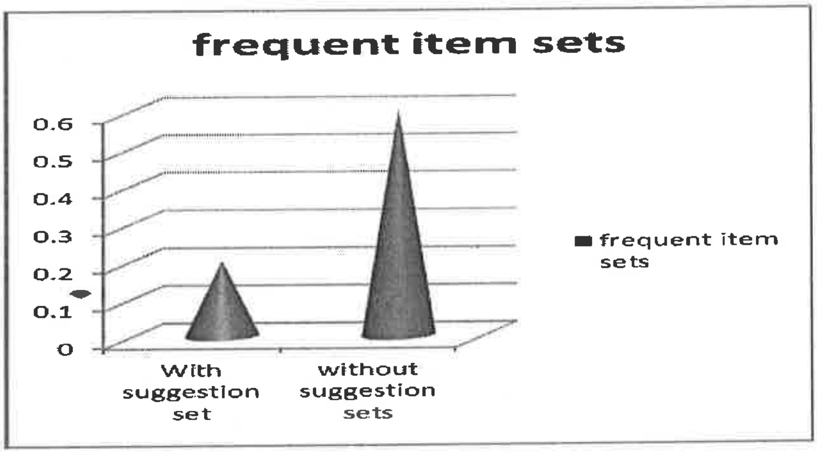
From the algorithm and its reasoning, one can see that the *FP-growth* mining process is a divide-and-conquer process, and the scale of shrinking is usually quite dramatic. If the shrinking factor is around 20-100 for constructing an FP-tree from a database, it is expected to be another hundreds of times reduction for constructing each conditional FP-tree from its already quite small conditional frequent pattern-base. Notice that even in the case that a database may generate an exponential number of frequent patterns, the size of the FP-tree is usually quite small and will never grow exponentially. For example, for a frequent pattern of length 100, "*a*1, . . . , *a*100", the FP-tree construction results in only one path of length 100 for it, possibly "_*a*1,→···→*a*100_" (if the items are ordered in the *list* of frequent items as *a*1, . . . , *a*100). The *FP-growth* algorithm will still generate about 1030 frequent patterns (if time permits!!), such as "*a*1, *a*2, . . ., *a*1*a*2,. . ., *a*1*a*2*a*3, . . ., *a*1 . . . *a*100." However, the FP-tree contains only one frequent pattern path of 100 nodes, and according to Lemma 3.2, there is even no need to construct any conditional FP-tree in order to find all the patterns.

V. RESULTS

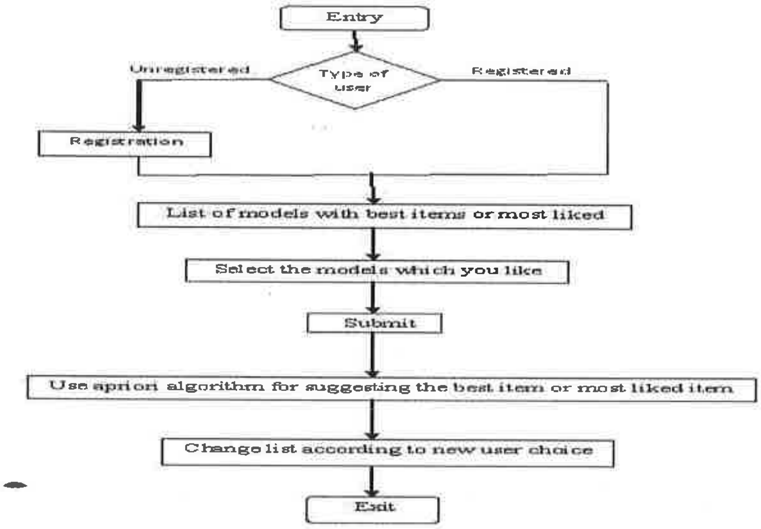
The above is the best method of ranking and suggesting the best methods in the scenario of mobile phone outlets in INDIA, which is shown in the following diagram:



As it was shown in the above diagram we were going to take the most liked items from the users and suggesting the best mobiles or the best set of suggestions that the most of the users liked or ordered.



The confidence of the suggestions were also proved by an traditional confidence calculations as follows In this section we are going to discuss about algorithms. Till now we have discussed some ranking rules , suggestion rules and Frequency move2set algorithm. We have some problems with these, so we go for an algorithm which suits our requirements well. The algorithm is Apriori algorithm. In order to know these algorithms we need to know some concepts of data mining.


Entry
Type of user
Unregistered
Registered
Registration
List of models with best items or most liked
Select the models which you like
Submit
Use apriori algorithm for suggesting the best item or most liked item
Change list according to new user choice
Exit

Frequent itemsets: Let I={I1, I2, I3,…., Im} be a set of items. Let D, the task-relevant data, be a set of database transactions where each transaction T is a set of items such that T is a subset of I. Each transaction is associated with an identifier, called TID. Let A be a set of items. A transaction T is said to contain A if and only if A is a subset of T. An association rule is an implication of the form A > B, where A is subset of I, B is subset of I and A∩B =Ø. The rule A > B holds in the transaction set D with support s, where s is the percentage of transactions in D that contain AUB. This is taken to be the probability ,P(AUB).The rule A > B has confidence c in the transaction set D, where c is the percentage of transactions in D containing A that also contain B. This is taken to be the conditional probability, P(B/A). That is, Support(A=>B) = P(AUB) Confidence(A=>B) = P(B/A) Rules that satisfy both a minimum support threshold (min_sup) and a minimum confidence threshold (min_conf) are called strong. The occurrence frequency of an itemset is the number of transactions that contain the itemset. This is also known, simply as the frequency, support count,or count of the itemset. The set of frequent k-itemset is commonly denoted by Lk. confidence(A > B) = P (A / B) = support(AUB) / support(A) = supportcount(AUB) / supportcount(A).

Mining frequent itemsets: In general, association rule mining can be viewed as a two-step process: 1. Finding all frequent itemsets: By definition, each of these itemsets will occur at least as frequently as a predetermined minimum support count, min-sup. 2. Generate strong association rules from the frequent itemsets: By definition, these rules must satisfy minimum support and minimum confidence

VI. CONCLUSION

All the previous process already proposed were very complex and contains very complicated computations which made the ranking and suggesting the best and popular items have been more and more complex and not getting to the actual end users. Now we have proposed as very simple randomized algorithm for ranking and suggesting popular items designed to account for popularity bias. This was utilized by many of the mobile outlets in the country successfully.

REFERENCES

[1]. Huidrom Romesh Chandra Singh, T. kalaikumaran, Dr. S. Karthik, Suggestion of True Popular Items, IJCSE, 2010.
[2]. Y.Maanasa, V.Kumar, P.Satish Babu, Framework for suggesting POPULAR ITEMS to users by Analyzing Randomized Algorithms, IJCTA, 2011.
[3]. V. Anantharam, P. Varaiya, and J. Walrand, —Asymptotically Efficient Allocation Rules for the Multiarmed Bandit Problem with Multiple Plays—Part i: i.i.d. Rewards,l IEEE Trans. Automatic Control, vol. 32, no. 11, pp. 968-976, Nov. 1987.
[4]. J.R. Anderson, —The Adaptive Nature of Human Categorizationl Psychological Rev., vol. 98, no. 3, pp. 409-429, 1991.
[5]. Yanbin Ye, Chia-Chu Chiang, A Parallel Apriori Algorithm for Frequent Itemsets Mining, IEEE, 2006.
[6]. Cong-Rui Ji, Zhi-Hong Deng, Mining Frequent Ordered Patterns without Candidate Generation.
[7]. Huang Chiung-Fen, Tsai Wen-Chih, Chen An-Pin, Application of new Apriori Algorithm MDNC to Exchange Traded Fund, International Conference on Computational Science and Engineering, 2009.
[8]. Milan Vojnovi_c, James Cruise, Dinan Gunawardena, and Peter Marbach, Ranking and Suggesting Popular Items, IEEE, 2009.

Improving Maximal Frequent Item set Mining for Sparse Dataset

B.sujatha Asst.prof, Ramesh babu Varugu Asst.prof

ABSTRACT

Mining of maximal frequent patterns is a basic problem in data mining applications. Small and useful association rules can be generated from maximal frequent itemset. The algorithms which are used to generate the maximal frequent patterns must perform efficiently. Most of the existing algorithms passed all frequent itemsets as candidates to the recursive algorithm which generates MFI. But the sparse dataset has huge number of frequent items and each frequent item has very small number of candidate items. This paper presents FastMFIMiner algorithm to generate MFI quickly from sparse dataset. It works efficiently even when the number of itemsets is more. The proposed algorithm has been compared with GenMax, Mafia and DepthProject for sparse and mushroom dataset and the results show that the proposed algorithm generates maximal frequent patterns quickly than existing algorithms.

Keywords: Maximal Frequent Patterns, Sparse Dataset, Mining MFIs

1. INTRODUCTION

A fundamental problem for mining association rules [5] is to mine frequent itemsets (FI's). In a transaction database, if we know the support of all frequent itemsets, the association rules generation is straightforward. However, when a transaction database contains large number of large frequent itemsets, mining all frequent itemsets might not be a good idea. The drawback of mining all frequent itemsets is that if there is a large frequent itemset with size n then almost all 2^n candidate subsets of the items might be generated. However, since frequent itemsets are upward closed, it is sufficient to discover only all maximal frequent itemsets (MFI's). Thus, a lot of work is focused on discovering only all the maximal frequent itemsets (MFIs).

There is much research on methods for generating all frequent itemsets efficiently [6, 7, 8] or just the set of maximal frequent itemsets [1, 2, 3, 4]. When the frequent patterns are long (more than 15 to 20 items), FI and even FCI become very large and most traditional methods count too many itemsets to be feasible. Straight Apriori-based algorithms count all of the 2 subsets of each k-itemset they discover, and thus do not scale for long itemsets. Other methods use “lookaheads” to reduce the number of itemsets to be counted. However, most of these algorithms use a breadth-first approach, i.e. Finding all k-itemsets before considering (k+1) itemsets. This approach limits the effectiveness of the lookaheads, since useful longer frequent patterns have not yet been discovered. Then, the merits of a depth-first approach have been recognized. The database representation is also an important factor in the efficiency of generating and counting itemsets.
Generating the itemset Z = (X ∪ Y) refers to creating t(Z) = t(X) ∩ t(Y), and counting is the process of determining support(Z) in T. Most previous algorithms use a horizontal row layout, with the database organized as a set of rows and each row representing a transaction. The alternative vertical column layout associates with each item X a set of transaction identifiers (tids) for the set t(X). The vertical representation allows simple and efficient support counting.

2. RELATED WORKS

Max Miner [1] is an algorithm introduced by Roberto Bayardo for finding the maximal frequent patterns. It uses efficient pruning techniques such as item reordering to quickly narrow the search. It introduces Support lower bound computation method for frequency computations. Max Miner employs a breadth first traversal of set enumeration tree of itemset. It reduces database scanning by employing a look ahead pruning strategy.

GenMax is a backtrack search based algorithm introduced by K. Gouda and M.J.Zaki [3] for mining maximal frequent itemsets. GenMax uses a vertical database format, where data is represented in item- tidset format. GenMax uses a number of optimizations to prune the search space. It introduces new techniques such as progressive focusing to perform fast superset checking, reordering for search space pruning and diffset propagation to perform fast frequency computation.

Depth Project [2] finds long itemsets using a depth first search of a lexicographic tree of itemsets, Depth project uses a bitstring representation of database and counting method based on transaction projections along its branches. Bucketing technique is used to improve the counting times. It returns superset of the MFI and requires post-pruning to eliminate non-maximal item sets.

Mafia [5] is one of the recent methods for mining the maximal frequent pattents. In Mafia the Search strategy combines a vertical bitmap representation of the database with an efficient relative bitmap compression schema. Mafia uses three pruning strategies to remove non-maximal sets. The first one is the look-ahead pruning introduced in MaxMiner. The second technique checks if t(X) ⊆ t(Y). If so X is considered together with Y for extension. The last method is to check if any existing maximal set includes the new set. Mafia requires a post-pruning step to eliminate non-maximal patterns.

3. PROPOSED WORK

When the MFI mining algorithms recursively construct many candidates, the performances of these approaches degraded, if the database is massive or the threshold for mining frequent patterns is low. Most of the existing algorithms take all frequent items as candidates from which all MFIs are generated. Instead of passing huge number of frequent items as candidates, here the recursive algorithm is invoked by every frequent item and its candidate pair because of frequent items having less number of candidates.

The main idea of the approach is to generate MFI quickly from the sparse dataset. In sparse dataset, there is lot of frequent itemset and each frequent itemset have less number of candidates. The maximal frequent itemset cardinality is not much smaller than frequent itemsets. The mean pattern length is also low. So instead of passing all frequent items as candidates, each frequent item and its candidates are passed to MineMFI algorithm. Tidset of frequent item is also sent, so that it is easy to compute the frequency of an itemsets. MineMFI algorithm is called for each frequent item and its candidate set pair. Once MineMFI is invoked, all MFI that include the particular frequent item are obtained.

The first step of FastMFIMiner is extracting all frequent items and reordering the frequent items in ascending order of their support. In second step, candidate items (CI_i) for each frequent item (FI_i) are generated and candidate items are arranged in increasing order of their
support. If $FI_i \cup CI_i$ has no superset in MFI and is frequent
then $FI_i \cup CI_i$ is added to MFI. The MineMFI
is called, if $FI_i \cup CI_i$ has no superset in MFI and $FI_i \cup CI_i$ is not frequent. The MineMFI algorithm is not called, if FIi ∪ CI_i is frequent or has superset in MFI. FastMFIMiner uses backtracking method to mine MFI and backtrack search space can be smaller than the full space because of using generating candidate and precede method. The FastMFIMiner generates candidates, once a frequent extension is obtained and generates maximal frequent itemset before finding all frequent itemsets.

The MineMFI method is invoked number of times which is less than or equal to the number of frequent item in the dataset. The FastMFIMiner method is not invoked, when the combination of FI_i and candidate items of FI_i is frequent. For sparse dataset this early finding of MFI has improved performance than other existing algorithms.

FastMFIMiner is explained with the following example. Let us consider the transaction database d which includes five different items, I = {A, B, C, D, E} and six transactions T= {1, 2, 3, 4, 5, 6}. The vertical data format of the database d is given in table 1. Support of an item is number of transactions that include the item.

All frequent items are extracted and reordered in ascending order with respect to the support. The support is directly given by the number of transactions in the tidset of each item. For example, consider the minimum support to be 3 transactions. In database d, all items are having more than two tids in the tidset, all items are frequent. The items A, B, C, D and E are reordered in ascending order with respect to the support and these are considered to next level. The frequent items are A, B, C, E and D.

Table 1. Vertical item -tidset of database D.

Item	Tidset
A	T1, T3, T5
B	T1,T3,T4,T5,T6
C	T1,T2,T3,T4,T5
D	T1,T2,T3,T4,T5,T6
E	T1,T2,T4,T5,T6

In the next level candidate items for each frequent item is obtained. All candidates are reordered in increasing order of support. The candidates of frequent item A is {B, C, D}, frequent item B is {C, E, D}, frequent item C is {E, D}, frequent item E is {D} and frequent item D is {}.

The first frequent item and its candidate set is A and {B, C, D} respectively. For this pair, FI ∪ CI has no superset in MFI and the itemset {A, B, C, D} is frequent and added to MFI. The next frequent item and its candidate set is B and {C, E, D} respectively. For this pair, FI ∪ CI has no superset in MFI and the itemset {B, C, D, E} is frequent and added to MFI. The subsequent frequent items and their candidate sets are C ∪ {E, D}, E ∪ {D} and D ∪ {}, having superset in MFI, ignored. MFI with support 3 transactions returned by FastMFIMiner are {A, B, C, D} and {B, C, D, E}.

Let us now consider the minimum support to be 4. In database d, all items except A are having more than three tids in the tidset, all items are frequent. The frequent items are B, C, D and E and are reordered in ascending order with respect to the support and these are considered to next level. Frequent items with support 4 in database d are B, C, E and D.

Candidate sets for each frequent item is generated in the next step. All candidates are reordered in increasing order of support. The candidates of frequent item B is {C, D, E}, frequent item C is {E, D}, frequent item E is {D} and frequent item D is {}.

The first frequent item and its candidate set is B and {C, E, D} respectively. For this pair, FI ∪ CI has no superset in MFI and the itemset{B, C, E, D} is not frequent. MineMFI method is called and the frequent item B, candidates of B (C, E, D) and tidset of B are passed to mine MFI which generates all MFI that includes the item B using backtracking method. MineMFI method returns the maximal frequent itemsets {B, D, E} and {B, C, D}.

The next frequent item and its candidate set is C and {E, D} respectively. For this pair, FI ∪ CI has no superset in MFI and the itemset {C, E, D} is frequent and is added to MFI.

The subsequent pairs are E ∪ {D} and D ∪ {} have superset in MFI and are ignored. MFI with support count 4 in database d [Table 1] are {B, C, D} and {B, D, E} and {C, E, D}.

MineMFI method is called only when FI_i ∪ CI_i is infrequent. The frequent item, candidate set and tidset(frequent item) is passed to MineMFI algorithm to mine MFI. MineMFI works as follows. MineMFI method gets three inputs. That are frequent item (FI_i), candidates of FI_i (CI_i) and Tidset(FI_i).

Definition1: FI denotes the set of frequent item and CI denotes the set of candidate items. FI_i denotes a frequent item and CI_i denotes the candidate items of frequent item FI_i. Each time the first element in candidate of FI_i (CI_i) is combined to FI_i and new candidates are generated for each frequent extension using generateCandidate method. Once the element is combined to FI_i, it is removed from candidates of FI_i (CI_i).

Definition2: Candidates of FI_i (CI_i) are generated from frequent items FI by removing FI_i. For example, frequent items are {A, B, C, D}. For frequent item A, the candidates are {B, C, D}. For frequent item B, the candidates are {C, D}. For frequent item C the candidates is {D} for frequent item D the candidates is {}.

The frequent itemset and its new candidate pair is added to MFI, if it has no superset in MFI and is frequent. Otherwise the next element in candidate of FI_i (CI_i) is combined to the FI_i and new candidates are generated.

Definition3: MineMFI method is not called, if FI_i ∪ CI_i is frequent and has no superset in MFI or it has superset in MFI.

Table 2. Maximal frequent itemsets.

Tid	Items	Maximal Frequent Itemsets Minimum support = 3	Maximal Frequent Itemsets Minimum support = 4
1	ABCDE	ABCD BCDE	BCD BDE CDE
2	CDE		
3	ABCD		
4	BCDE		
5	ABCDE		
6	BDE		

The maximal frequent itemsets of database d[Table 1] for support 4 and 3 transaction is given in Table 2. MFI miner mines maximal frequent itemsets using depth first search strategy, generates candidates whenever a frequent itemset is generated. This process is repeated until there is no candidate for frequent itemset. In case of sparse datasets, data contains small patterns and there is not much repetition of patterns in dataset. FastMFIMiner has better performance than the existing algorithms.

Algorithm – FastMFIMiner

Input: dataset D, support S

```
Output: Maximal Frequent Itemsets MFI
FastMFIMiner (Dataset D, Support S) BEGIN
    1. Generate frequent items and reorder them in
       ascending order of their support.
    2. Generate candidate items for each frequent item
       and reorder the candidates in increasing order of the
       support.
    3. For each x ε FIs
         a. If x ∪ candidates(x) has no superset in
            MFI
              i. if size of x is 1 or 2 then add x to
                 MFI
            // most of the sparse dataset the candidate
            items of frequent item may be 1 or  2;
             ii. else if x ∪ candidates(x) is
                 frequent then add x to MFI;
            iii. else MineMFI(x,
                 candidates(x), tid(x))
END
```

```
MineMFI
Input: frequent item, candidate set, Tidset of the
frequent item.
Output: Maximal Frequent Itemsets that includes the
frequent item.
MineMFI (frequent item, candidateset, ftids)
BEGIN
    For each x ε candidateset
        If frequent ∪ candidateset has
    superset in MFI then return;
        Nfrequent = frequent ∪ x    // current
    frequent itemset
        Ntids=x.tid ∩ ftids      // Tidset of current
    frequent itemset
        candidateset.remove(x)   // candidates of
    current frequent itemset
        if candidateset is empty &&
    Nfrequent has no superset in MFI
            MFI.add(Nfrequent); return;
        newcandidate =
    generatecandidates(Nfrequent,candidateset,Nti ds)
        //exact canididate of current frequent ietmset
        if newcandidate is empty
        if Nfrequent has no superset in MFI
            MFI.add(Nfrequent);
        else
        generateMFI (Nfrequent,
    newcandidate,Ntids)
END
```

```
Generate Candidates
Input: Frequent Itemset, Candidate set, Tidset of the
Frequent Itemset.
Output: Exact Candidates of Frequent Itemset.
Generate Candidates (frequent, candidate, ftids)
BEGIN
    cand=null; // cand will contain the exact
        candidates of frequent item(frequent)
    for each x ε candidate
        If(ftids ∩ tid(x) ≥ support)
        cand.add(x); // candidates are stored in
    increasing order of support.
    return (cand);
END
```

The MineMFI method follows depth first search strategy and candidate generation method. Using depth first search maximal frequent itemsets can be mined before generating all frequent Itemsets. In breath first search method, all I_{l+1} - frequent itemsets are obtained, after obtaining all I_l - frequent itemsets. Tidsets of each frequent itemset is obtained and passed to the candidategenerate method. Once the tidset of a frequent itemset is obtained, the candidates of frequent extensions are obtained easily from the tidset of frequent itemset. This process reduces the frequency computation time.

Pruning

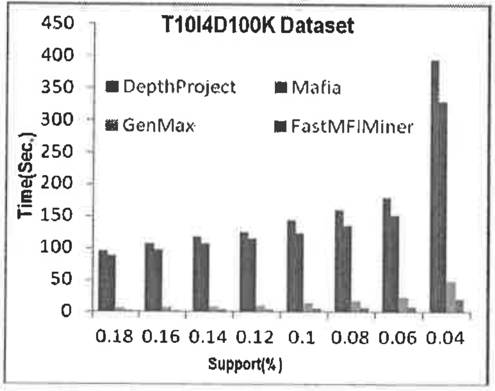
Most of the standard algorithm like mafia [4], depthproject[2], genmax[3] get all frequent items as candidates and MFIs are obtained from these candidates. To generate the MFI quickly, FastMFIMiner uses two pruning techniques. The first pruning technique is reordering of items with respect to its support in ascending order. This reordering technique is introduced by Roberto Bayardo in MaxMiner algorithm [1] for mining maximal frequent itemsets. Once frequent itemse are generated, they are reordered with respect to its support in ascending order. The second pruning technique is the recursive method MineMFI is not called, if the combination of frequent item (FIi) and candidates of FI_i (CI_i) (FIi ∪ CIi) is frequent or FIi ∪ CIi has superset in MFI.

Performance Evaluation

The performance of FastMFIMiner algorithm is compared with three different algorithms, it is observed that the performance is varied significantly depending on the dataset characteristics. To evaluate the performance of FastMFIMiner algorithm, four different benchmark datasets is used. All these datasets are downloaded from FIMI Repository [9]. Dataset taken for experiment are T10I4D100K, T40I10D100K, Retail and Mushroom Dataset.

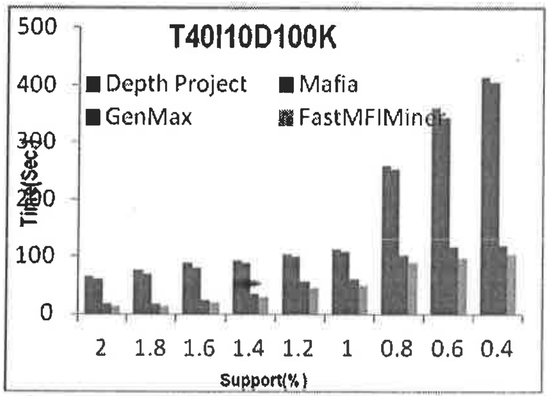
The first dataset is T10I4D100K which contains 1000 attributes and 100,000 records. The average record length is 10. The mean pattern length is very small and it is around 2 to 3. In T10I4D100K dataset, the number of frequent items is huge and frequent itemset will have small number of candidates. The performance of FastMFIMiner algorithm has been compared with

GenMax[3], Mafia[4], and Depth Project[2] algorithm for various support and results show that FastMFIMiner gives superior performance than the existing algorithms. Figure 1 illustrates that, the FastMFIMiner algorithm has better performance, when compared to conventional GenMax, Mafia and Depth Project algorithm.



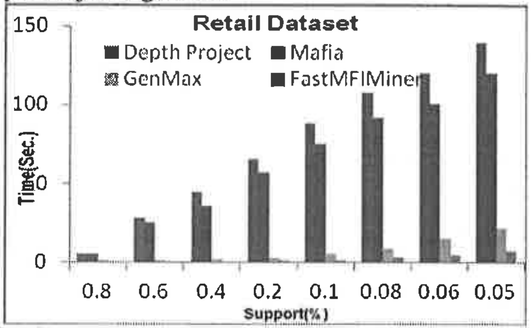
Figs.1: Performance of FastMFIMiner algorithm on T10I4D100K dataset.

The second dataset is T40I10D100K which contains 1000 attributes and 100,000 records. The average record length is 40. In T40I10D100K dataset, the number of frequent items is huge and frequent itemset will have small number of candidates. The mean pattern length is very small and it is around 2 to 6. Mean pattern length is slightly greater than T10I4D100K dataset. Number of maximal frequent itemset is not much smaller than the number of all frequent patterns. Figure 2 illustrates that, the FastMFIMiner algorithm has better performance than GenMax, Mafia and Depth Project algorithms on T40I10D100K dataset.



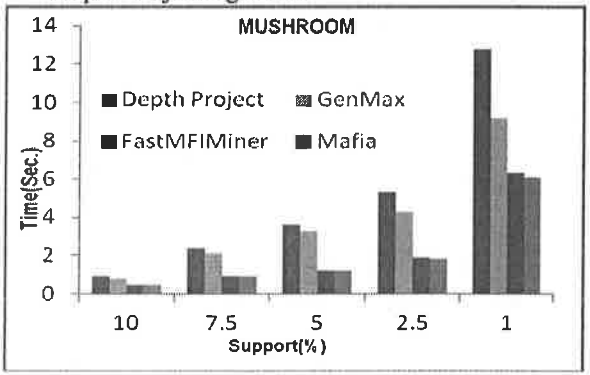
Figs.3 Performance of FastMFIMiner algorithm on T40I10D100K dataset.

The third dataset is Retail which contains 16,470 items and 88,162 transaction. The average number of distinct items in each transaction is 13 and most transaction contains items between 7 and 11 items. Mean pattern length is also very short. The maximum number of maximal patterns is of length one or two. Number of maximal frequent itemsets is not much smaller than the number of all frequent itemsets. In retail dataset, the number of frequent items is huge and frequent itemset will have small number of candidates. Figure 3 illustrates that, the FastMFIMiner aalgorithm has better performance than GenMax and Mafia and Depth Project algorithms on Retail dataset



Figs.3 performance of FastMFIMiner algorithm on Retail dataset.

Mushroom dataset contains 120 items and 8,124 transactions. The average transaction length for mushroom is 23 thus a maximal pattern spans almost a full transaction. The total number of maximal frequent itemset is about 1000 times smaller than all frequent itemsets [3]. On mushroom dataset, the results of FastMFIMiner algorithm are similar to Mafia. FastMFIMiner algorithm has better performance than GenMax. Because in mushroom dataset a smaller itemset have many number of maximal frequent itemsets. To perform maximality checking one has to test against a large set of maximal itemsets. GenMax has better performance than Depthproject. Figure 4 illustrates that, the FastMFIMiner algorithm has better performance than GenMax and Depth Project algorithms on Mushroom dataset.



Figs. 4 performance of FastMFIMiner algorithm on Mushroom dataset.

4. CONCLUSION

In this paper we have introduced FastMFIMiner algorithm for mining maximal frequent patterns quickly from sparse dataset. The initial candidates for every frequent item are generated by finding association among frequent items. The generated candidates are sorted in increasing order of their support. The frequent item ∪ its candidate set are added to MFI directly, if the $FI_i \cup CI_i$ is frequent and has no superset in MFI. Otherwise MineMFI is invoked to obtain MFI from frequent items FI_i and candidate set CI_i. The performance of FastMFIMiner is tested using benchmark dataset and results show that, FastMFIMiner generates MFI very quickly from sparse dataset.

REFERENCES

[1] Bayardo, R.J., 1998. Efficiently mining longpatterns from databases. In the Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data, Seattle, June 001-04, Washington, United States, pp: 85-93. http://doi.acm.org/10.1145/276304.27631

[2] Agrawal, R., C. Aggarwal and V. Prasad, 2000. Depth first generation of long patterns. In the Proceedings of the 6th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining, Aug. 20-23, Boston, Massachusetts, United States, pp: 108-118. http://doi.acm.org/10.1145/347090.347114

[3] Gouda, K. and M.J. Zaki, 2001. Efficiently mining maximal frequent itemsets. In the Proceedings of International Conference on Data Mining, Nov. 29-Dec. 02 2001, IEEE Computer Society Washington, DC, USA., pp:

163-170. http://portal.acm.org/citation. cfm?id = 645496.6580 47&coll=GUIDE&dl=GUIDE

[4] Burdick, D., M. Calimlim and J. Gehrke, 2001. MAFIA: A maximal frequent itemset algorithm for transactional databases. In International Conference on Data Engineering, Apr. 02-06, IEEE Computer Society Washington, DC, USA pp:443-452.

[5] Agrawal, R. and R. Srikant, 1994. Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Databases, Sep. 12-15, Santiago de chile, Chile, pp: 487-499. DOI: 10.1.1.40.7506

[6] Savasere, A.,E. Omiecinski and S. Navathe, 1995.An efficient algorithm for mining association rules in large databases. Proceedings of 21 International VLDB Conference on Very Large Data Bases, Sep.
11-15, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA ., pp:432-444. http://portal/acm.org/citation.cfm?id=673300

[7] Ramesh C. Agarwal, Charu C. Aggarwal and V.V.V. Prasad,2001. A tree projection algorithm for generation of frequent itemsets. J. Parallel Distribut. Comput., 61:
350-371. DOI: 10.1006/ jpdc.2000.1693

[8] R. Agrawal and R. Srikant. Fast algorithms for mining association rules.In J.B. Bocca, M. Jarke, and C. Zaniolo, editors, Proceedings
20th International Conference on Very Large Data Bases, pages 487–499. Morgan Kaufmann, 1994. [9].Lin, D.I. and Z.M. Kedem, 1998. Pincer search: A new algorithm for discovering the maximum frequent sets. In Proceedings of the
6thInternational Conference on Extending Database Technology, Mar. 23-27, Springer- Verlag London,UK.,pp:105-119
http://portal.acm.org/citation.cfm?
id=645338.6503 96.

[9] FIMI 2004- Frequent itemset Mining Implementations URLs
i) http://www.adrem.ua.ac.be/~goethals/
ii) http://fimi.cs.helsinki.fi/fimi04/
iii)http:/ icdm-04.cs.uni-dortmund.de/

International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 3, Issue 3, May – June 2014 **ISSN 2278-6856**

Enterprise Resource Planning – Analysis of Business Intelligence & Emergence of Mining Objects

Ramesh Babu Varugu

Asst Prof in Ananamacharya Institute of science & Technology

Abstract: ***Build a model to investigate system and discovering relations that connect variables in a database are the tasks of data mining when concise valuable knowledge about the system of interest discovered and should be incorporated into some decision support system which helps the manager to make business decisions. Operational reports from the ERP system do not satisfy mangers requirements for planned versus actual monitoring forecasting exception analysis without business intelligence users must compile these reports manually from standard report. Business intelligence system can unlock the value of the data in ERP reports and daily order report that will help in better decision making and can be of use to the decision makes in more than one department and the system provide online analytical processing and data mining tools that user can use to answer many question to obtain progressively more detailed information about retain sales change metrics view graphs and charts reuse reports. This paper presents the Enterprise Resource Planning survey and modules used in applications that are effective in business, the modules such as Business Objects & Business warehouse are decision making task when we execute the report based the on the output the person person can identify their loss or profit.***

Keywords – Enterprise Resource Planning, Business Objects, Business Warehouse, Mining, SRM & CRM.

1. INTRODUCTION

Enterprise resource planning extended beyond the boundaries of an organization to integrate with the information systems of other stakeholders and partners like customers and suppliers, standard ERP system and to further enhance its capabilities. Data warehousing is a concept information system in an organization by product of any business transaction which keeps on growing over period of time. Entire operational data is kept in the database of the ERP system it will affect the performance of the ERP system. Thus an organization after the use of this operational data is over would like to remove this data and store in other folder it in some large data store. Data warehouse applications are important as relieve the main ERP system from the tasks of data extraction cleaning and loading the data onto the system and thus can avoid unnecessary burden on the main system for example to analyse the sale data store for ten years in transaction entries for that period data warehousing come in to picture, if we not use the warehouse concept the system might go slow.

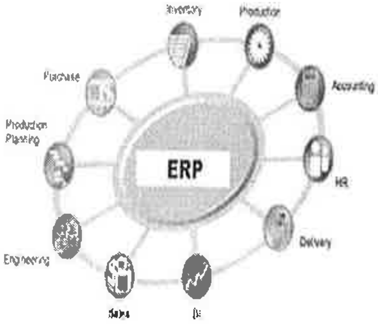


Figure 1 ERP Application Modules

Enterprise resource planning is a way to integrate the data and processes of an organization into one single system includes hardware and software in order to achieve integration, most ERP systems use a unified database to store data for various functions found throughout the organization. The term ERP originally referred to how a large organization planned to use wide resources were used in larger more industrial types of companies, today the term can refer to any type of company no matter what industry it falls in. For a software system to be considered ERP it must provide an organization with functionality for two or more systems while some ERP packages exist that only exist that only cover two functions for an organization. ERP systems can cover wide range of functions and integrate them into one unified database for instance functions such as human resource supply chain management financials etc. Traditional decision support system when responsible personnel are in the decision making process the information provided by traditional decision support systems may be inadequate information may not have reached optimal levels of effectiveness. Enterprise resource planning data mining online analytical processing artificial intelligence the massive amounts of data assets accumulated

SECTION II

2. Related Work: Enterprise resource planning information system designed to coordinate all the resources information and activities needed to complete business processes such as order billing. ERP supports most of the business system that maintains in a single database the data needed for a variety of business functions such as Manufacturing, Supply Chain

Management, Financials, Projects Human Resources and Customer Relationship Management. An ERP system based on a common database and modular software design, it can each department of a business to store and retrieve information in real-time. The information should be reliable accessible and easily shared. The data for various business functions are integrated, ERP system may comprise a set of discrete applications each maintaining a discrete data store within one physical database. The term ERP originally referred to a large organization planned to use organizational wide resources. ERP systems were used in large industrial types of companies however the use of ERP has changed and is extremely comprehensive. ERP systems can cover a wide range of functions and integrate them into one unified database functions like Human Resources, Supply Chain Management, Customer Relations Management, Financials, Manufacturing functions and Warehouse Management functions were all once stand- alone software applications. Data warehousing is a repository of an organizations stored data can be accessed by an organizations customers suppliers and employees designed to facilitate for reporting and analysis. Data warehouse focuses on data storage means retrieve and analyse data to extract transform and load data and to manage the data dictionary are also considered essential components of a data warehousing system. Implementing an ERP system not an easy task to achieve takes lots of planning consulting extends one year, wide range of scope and for many larger organizations can be extremely complex. ERP system will ultimately require significant changes on staff and work practices. One of the most important traits that an organization should have when implementing an ERP system is ownership of the project. Because so many changes take place and its broad effect on almost every individual in the organization, it is important to make sure that everyone is on board and will help make the project and using the new ERP system a success. Usually organizations use ERP vendors or consulting companies to implement their customized ERP system. There are three types of professional services that are provided when implementing an ERP system, they are Consulting, Customization and Support. Consulting Services – usually consulting services are responsible for the initial stages of ERP implementation, they help an organization go live with their new system, with product training, workflow, improve ERP's use in the specific organization, etc. Customization services work by extending the use of the new ERP system or changing its use by creating customized interfaces and/or underlying application code. While ERP systems are made for many core routines, there are still some needs that need to be built or customized for an organization. Support Services- Support services include both support and maintenance of ERP systems. For instance, trouble shooting and assistance with ERP issues. Decision systems are the combination of information technology and artificial intelligence such systems allow organizational networks to work toward the computerization of management systems which are made up of particular regulations and tasks. Performing all types of responses and management decision making the computerized systems will permit fast cost effective responses to unpredictable and ever changing product demand and support rapid product launches for previously unplanned products tailored to meet changing customer desires. Organizational data mining is defined as leveraging data mining tools and technologies to enhance the decision making process by transforming data into valuable and actionable knowledge to gain a competitive advantage.

SECTION III

3. Enterprise Resource Planning: ERP is a software module includes for finance manufacturing production planning human resources plant maintenance material management quality management marketing sales and distribution order tracking finance accounting marketing and HR.

Finance: Finance module is the core of many ERP software systems gathers finance data from various functional departments and generates valuable financial reports such as the balance sheet general ledger trail balance and quarterly finance statements.

Manufacturing: Enables an enterprise to many technologies with business processes to create an integrated solution it provides the information based upon which the entire operation should be run. It provides the freedom to change manufacturing and planning methods as needed.

Human Resource: Widely implemented in Enterprise resource planning streamlines the management of human resources and human capitals routinely maintain a complete employee database including contact information salary details attendance performance evaluation and promotion of all employees.

Materials Management: Maintaining the appropriate level of stock in warehousing integration of the inventory control module with sales purchase and finance module allows ERP systems to generate vigilant executive level reports.

Production planning: optimizes the utilization of manufacturing capacity parts components and materials resources using historical production data and sales forecasting.

Plant Maintenance: information of this module helps to reduce the duration and cost of plant downtime as a result of damage and to recognize possible weak points within technical system in good time.

Sales and Distribution: Provides a complete sales management solution for a broad range of industries and part of the logistics module that support customers starting from quotations sales order and all the way towards billing the customer integrated with Materials Management and Production Plan

Supply Chain Management: Supply chains perform the companies and the business activities needed to design make deliver and use a product or service. Business depends on their supply chains to provide them with what they need to survive and thrive. The pace of change and the uncertainty about how markets will evolve has made it increasingly important for companies to be aware of the supply chains they participate in and to understand the roles that they play. Term "Supply Chain Management" as logistics and operations management is the alignment of firms that bring products or services to market supply chain consists of all stages involved directly or indirectly in fulfilling a customer request. Supply chain is a network of facilities and distribution options that performs the functions of procurement of materials transformation of these materials into intermediate and finished products to customers. Supply chain management is the coordination of production inventory location and transportation among the participants in a supply chain to achieve the best mix of responsiveness and efficiency for the market being served.

Customer Relationship Management: Customer relationship management entails all aspects of interaction that a company has with its customer whether it is sales or service related, describes a business customer relationship. Customer relationship management solutions provide with the customer business data to help you provide services or products that customers want provide better service cross sell and up sell more effectively. Most business realize when moving to a CRM system comes directly from having all business data stores and accessed from a single location, before CRM systems customer data was spread out over office productivity suite documents emails systems mobile phone data even paper note cards and rolodex.

Helping an enterprise to enable its marketing departments to identify and target their best customers manage marketing campaigns and generate quality leads for the sales team.

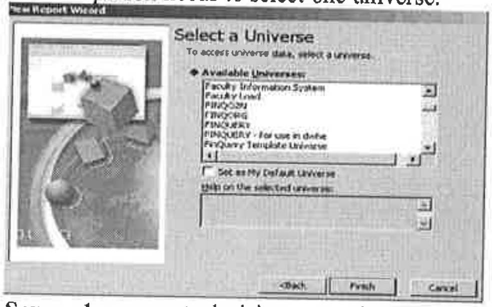
Assisting the organization to improve telesales account sales management by optimizing information shared by multiple employees and streamlining existing processes

Allowing the formation of individualized relationships with customer with the aim of improving customer satisfaction and maximizing profits identifying the most profitable customers and providing them the highest level of service.

Providing employees with the information and processes necessary to know their customer understand and identify customer needs and effectively build relationships between the companies its customer base and distribution partners.

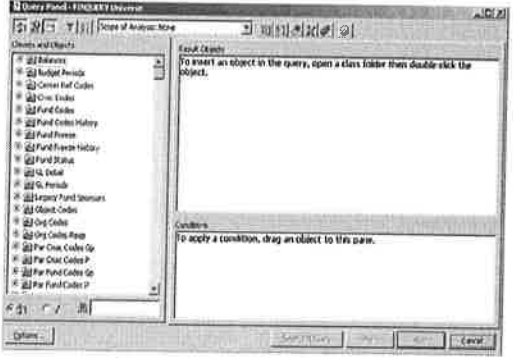
SECTION IV

4.1. Enterprise Resource Planning Modules: Business objects is a collection of classes that is intended for a particular group of users to meet the needs of persons who access and collecting specific data in the warehouse. Created one academic example for the purpose of student enquiry below screen shows the report to access universe data that the person needs to select one universe.


New Report Wizard
Select a Universe
To access universe data, select a universe.
Available Universes:
Faculty Information System
Faculty Load
FINQS2N
FINQORG
FINQUERY
FINQUERY - For use in dwhe
FinQuery Template Universe
Set as My Default Universe
Help on the selected universe:
<Back
Finish
Cancel

Screen 1 represents decision on preferred course

Class is a logical collection of objects, classes in business object correspond to a table in the data warehouse. The above example shows the student enquiry business object corresponds to the degree received table in the warehouse just as the balance class in the universe corresponds to the balance table in the warehouse. The related warehouse table degree received class contains objects that provide data about degrees received at universe and other institutions. Some classes can contain objects that will return data from more than one table though this is infrequent in the current universe.

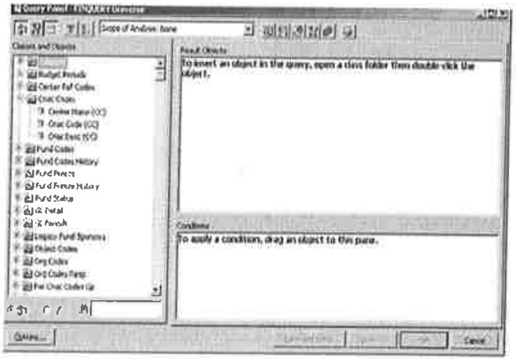


To insert an object in the query, open a class folder then double click the object.

To apply a condition, drag an object to this pane.

Screen 2 is a class for report

An object corresponds to a data element a calculation or a function based on one or more data elements selected to construct a query on the warehouse. Once the query is performed the objects are returned with corresponding values.



To insert an object in the query, open a class folder then double click the object.

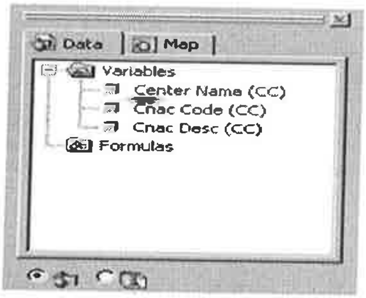
To apply a condition, drag an object to this pane.

Screen 3 is object for report

Dimension objects retrieve the data that will provide the basis for analysis in a report, it typically retrieve the character type data or accounting periods etc. Measure objects retrieve numeric data that is the result calculation on data in the database, objects are semantically dynamic the values they return depend on the objects they are used which include Accounting period actual month in a query from the universe, business object calculates the actual month balance per object for that organization in that accounting period.

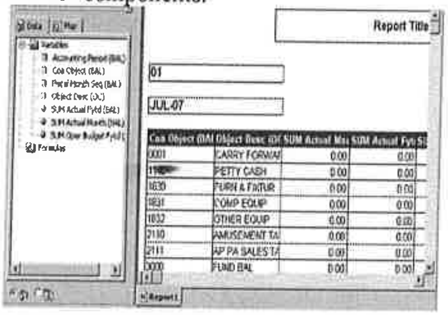
To retrieve the data from warehouse need to give a query, aim of building query is to retrieve the data that you want to view and analyze in a report. Provides the query panel the one stop graphical interface in which we perform some tasks like view the classes and objects of the universe, select the objects we want to include in the query and run the query.

Report manager is a key of the desktop intelligence workspace that enables to manage many different aspects of work quickly and easily from one window can manage all the variables in one report work on the structure and formatting of report components and use the navigation view to go quickly from report to report. When the report manger window opens it is docked on the left hand side of report of report window, we can also undock the report manager window and drag it to any other convenient location on screen and hold down the CtrlKey while moving the report manger window to prevent it from docking.



Screen 4 Business object tools consist of two tabs one is Data and Map.

Data tab is used to manage the variables in report it contains a list of the objects, variables and formula in a document even we can drag items from this list and drop them directly into the report window to construct or edit tables and other components.



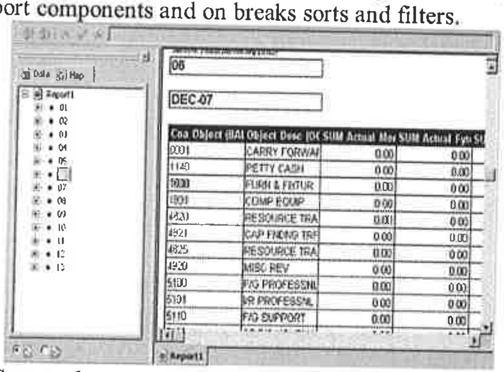
01

JUL-07

Coa Object (BAL	Object Desc (OC	SUM Actual Mo	SUM Actual Fyt
0001	CARRY FORWAI	0.00	0.00
[illegible]	PETTY CASH	0.00	0.00
1830	FURN & FIXTUR	0.00	0.00
1831	COMP EQUIP	0.00	0.00
1832	OTHER EQUIP	0.00	0.00
2110	AMUSEMENT TA	0.00	0.00
2111	AP PA SALES TA	0.00	0.00
3000	FUND BAL	0.00	0.00

Screen 5 represents how the data stored in Data tab.

Map tab contains navigation view displays a list of all the reports in the desktop intelligence document currently displayed on the screen and structure view displays a list of all the components in the selected report each report component is represented by an icon and a name hidden in the report are displayed in italics when click on an icon in the report manager window the corresponding component is displayed in the main report window. We can drag and rearrange the components in the report from one position to other position, formatting of the different report components and on breaks sorts and filters.



06

DEC-07

Coa Object (BAL	Object Desc (OC	SUM Actual Mo	SUM Actual Fyt
0001	CARRY FORWAI	0.00	0.00
1140	PETTY CASH	0.00	0.00
1830	FURN & FIXTUR	0.00	0.00
1831	COMP EQUIP	0.00	0.00
4920	RESOURCE TRA	0.00	0.00
4921	CAP FNDNG TRF	0.00	0.00
4825	RESOURCE TRA	0.00	0.00
4930	MISC REV	0.00	0.00
5100	F/G PROFESSNL	0.00	0.00
5101	VR PROFESSNL	0.00	0.00
5110	F/G SUPPORT	0.00	0.00

Screen 6 represents how the data stored in Map tab.

4.2. Business Warehouse: Business Warehouse integrates transform and consolidates relevant business information from productive applications and external data sources which provides high performance infrastructure that helps to evaluate and interpret data. Data warehousing in NetWeaver business warehouse enables the integration transformation consolidation clean up and storage of data incorporates the extraction of data for analysis and interpretation. Data warehousing process includes data modelling, data extraction further processing of data and the administration of the data warehouse management processes. Analytic engine provides OLAP functions and services as well as services for business warehouse integrated planning and analysis process design. This tool include query reporting and analysis functions as an employee with access authorization can evaluate past current data on various levels of detail and from different perspectives on the web and Microsoft excel. Business tool create planning applications and data entry in business warehouse integrated planning, information broadcast to distribute content from business intelligence by email pre-calculated documents with historical data or as links with live data. NetWeaver business warehouse support in developing supplying data in the system landscape performing tests and traces as well as monitoring business warehouse in system landscape, business warehouse provides an open architecture in many areas can extract data from various sources and load the data into a business warehouse system and evaluate this data for reporting using various front end tools. Business warehouse enables highly efficient processing of queries and data warehouse load processes.

Business warehouse planning users generally need the authorization as for analysis displaying and to change the data, the planning provides business experts with an infrastructure for creating and operating planning scenarios or other applications. Creating business data requires data store objects for direct updates in planning mode are used to store data. It contains aggregation levels that can be entered or changed manually by user input automatically by a planning function consists of subset of the characteristics and key figures of a multi provider for direct updates in planning mode.

SECTION V

5.1. Problem Definition: Enterprise resource planning solution oftentimes makes it difficult to adapt to the specific needs of individual organizations because ERP software is enormously sophisticated there is often a tendency to implement more features and functions for a particular installation than is actually needed. Cost to implement and maintain ERP systems is very high and some departments and users may be hesitant to agree to the implementation if they feel they are giving up control of their data. Large industrial companies installed enterprise resource planning systems that massive computer applications allowing a business to manage all of its operations like functions requirements planning, human resources and databases on the basis of a single integrated set of corporate data.

5.2. Mining ERP Application: Business objects Business warehouse & intelligence is a reporting tool dynamically restricts data being returned by a query. Hindustan unilever is a Sales & Distribution Company, all the data maintains in enterprise resource planning that is systems applications & products. The functionality of business intelligence is when the Hindustan unilever person executes the reports in business intelligence server, the output of business intelligence gets the decision chats, pie charts and graphs. Here the functionality of business intelligence is based on the graph or pie chart person identifies how the products sales profit or loss going in there outlets.

6. CONCLUSION

In this paper we have outlined mining tools in ERP efforts to address important and challenging issues of accuracy efficiency and usability of business intelligence data. Mining tool on Enterprise resource planning how it was used in decision making scenarios on sales and distribution inventory etc. Our extend with one implementation on Business Intelligence report with real-time example.

Reference

[1] Alex Berson Stephen Smith and Kurt Thearling "Building Data Mining Applications for CRM" McGraw-Hill 2000.

[2] Adomavicius, G., and Tuzhilin, A. 2005. "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Transactions on Knowledge and Data Engineering* (17:6), pp. 734-749.

[3] Associated Press. 2012. "Columbia U Plans New Institute for Data Sciences," July 30 (http://www.cbsnews.com/8301-505245_162-57482466/columbia-u-plans-new-institute-for-data-sciences/, accessed August 3, 2012).

[4] Daniel E O'Leary "Enterprise Resource Planning Systems Life Cycle Electronic Commerce and Risk Cambridge University Press 2000.

AUTHORS:

Ramesh Babu Varugu B.Tech Computer Science Engineering from Lakkireddy Balreddy Engineering College M.Tech Information Technology from Gurunank Engineering College. Having eight years of experience in Academic currently working as Asst Prof in Ananamacharya Institute of science & Technology. He has guided many UG & PG students interested subjects include Network Security, Data Mining, Cloud Computing.

IJEEE
International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 1, Issue 8, August 2014)

Tree Based Graph Mining – Analysis of Interaction Pattern Discovery in Business

Ramesh Babu Varugu
Computer Science Dept.
Ananamacharya Institute of science & Technology

C. Shanker
Computer Science Dept.
Sri Indu College of Engineering & Technology

***Abstract*— To Discussion business requirement integral part of organization also an important communication and coordinate activity of teams status are discussed decisions ideas are generated. Data mining technique to detect frequent interaction patterns and extraction of pattern mining algorithm was proposed to analysis tree structure and retrieve flow patterns, we present the analysis of tree. Mining algorithm extracts the frequent patterns of human interaction explore tree mining for hidden interaction pattern discovery using classification of mining algorithms from captured content of time series of many meetings in particular time periods years ago. Our system provides the efficient method extract the information priority wise algorithms compare to earlier technique.**

***Index Terms*— Human Interaction, Mining Algorithms, Tree structure Graph, Human Interaction.**

1. INTRODUCTION

The data to be mined is first extracted from an enterprise data warehouse into a data mining database or data mart, cleaning data for a data warehouse and for data mining are very similar once the data has been cleansed for a data warehouse then it most likely will not need further cleansing in order to be mined. The data mining database may be a logical rather than a physical subset of data warehouse provided that the warehouse can support the additional resource demands of data mining. Scripting engine was designed and developed for the data extraction layer to pull reports either manually or as a scheduled task from the data warehouse repositories created by the claims examiners are stored in Microsoft by the metadata repository.

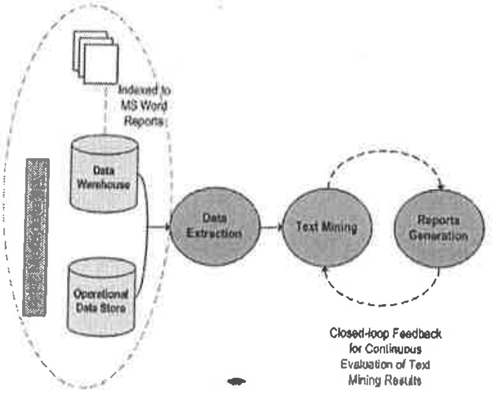


Figure 1 Represents the Mining Text data from Data Warehouse

Mining scores based on historically analysis of likelihood of fraud were custom developed based on text entered in the claims reports and details based on the claims. Reports are generated on a web based application layer fed into the enterprise resource planning application which is used by the client and commonly found in most of the larger insurance companies. Many text mining applications give users open ended freedom to explore text for meaning. Text mining can be used as a deeper more penetrative method which goes beyond escalations of possible search interests and to sense the mood of the written text. Text mining is the method that supports users to find useful information from a large amount of a digital text data should retrieve the information that users require with relevant efficiency. Information Retrieval has the objective of automatically retrieving as many relevant documents as possible filtering out irrelevant documents at the same time. However Information retrieval based system do not adequately provide users with what really need. Many text mining methods have been developed in order to achieve the goal of retrieving for information users, process of extracting discovery pattern consist of following Data Selection, Data Processing, Data Transaction, Pattern Discovery & Pattern Evaluation. The ability to search for keywords in a collection is widespread such search only marginally supports discovery because the user has to decide and can suggest interesting patterns to look at and the user can then accept or reject these pattern as interesting.

Data mining, which is a powerful method of discovering new knowledge, has been widely adopted in many fields, such as bioinformatics, marketing, and security. In this study, we investigate data mining techniques to detect and analyze frequent interaction patterns; we hope to discover various types of new knowledge on interactions. Human interaction flow in a discussion session is represented as a tree. Inspired by tree-based mining, we designed interaction tree pattern mining algorithms to analyze tree structures and extract interaction flow patterns. An interaction flow that appears frequently reveals relationships between different types of interactions. Mining human interactions is important for accessing and understanding meeting content. First, the mining results can be used for indexing meeting semantics, also existing meeting capture systems could use this technique as a smarter indexing tool to search and access particular semantics of the meetings. Second, the extracted patterns are useful for interpreting human interaction in meetings. Cognitive science researchers could use them as domain knowledge for further analysis of human interaction.

Moreover, the discovered patterns can be utilized to evaluate whether a meeting discussion is efficient and to compare two meeting discussions using interaction flow as a key feature.

2. RELATED WORK

Text mining involves a large collection of unrelated digital items in a systematic way and to discover previously unknown facts which might take the form of relationships or patterns that are build deep in an extensive collection. Information retrieval system identify the documents in a collection which match a user query, search engine which allows identification of a set of documents that relate to a set of key words. Information extraction is the process of automatically obtaining structured data from an unstructured natural language document. Term analysis identifies the terms in a document where a term may consist of one or more words especially useful for documents that contain many complex muti word terms such as scientific. Named-entity recognition identifies the names in a document such as the names of people systems are also able to recognize dates and expressions of time quantities and associated units percentages and fact extraction which identifies and extracts complex facts from documents such facts could be relationship between entities or events. Data mining is the process of identifying patterns in large sets of data when used in text mining is applied to the facts generated by the information extraction phase and the result of data mining process are put into another database that can be queried by the end-user via a suitable graphical interface network of protein interactions.

Electronic discovery refers to discovery deals with the exchange of information in electronic format and agreed upon processes and often reviewed for privilege and relevance before being turned over to opposing. Data are identified as relevant and extracted using digital forensic procedures and is reviewed using a document review and useful for its ability to aggregate. Electronic information is different from paper information because of its intangible form usually accompanied by metadata that is not found and can play an important part as evidence. Providing a text mining for science requires a new means of collaboration between existing and future stakeholders to accept data and text mining as being effective and acceptable processes such mining does not eliminate any significant role currently being performed by stakeholders that it does not raise challenges and barriers to text mining applications that it does not threaten publishers. Text mining is believed to have a considerable commercial value particularly true in scientific disciplines in which highly relevant information is often contained within written text.
A distributed model raises issues around data normalization of performance levels of other standardization issues requires conformity by all involved to common metadata standards to allow effective cross reference and indexing. In support of text mining one can see the emergence of the cloud as a mechanism for processing large amounts of data using the existing powerful computer resources made available by organizations such as Amazon, Microsoft, HP, etc.

3. ALGORITHMS USED

Purpose of the analysis in some instances the extraction of semantic dimensions alone can be useful outcome if it clarifies the underlying structure of input documents. To reiterate text mining can be summarized as a process of numericizing text.

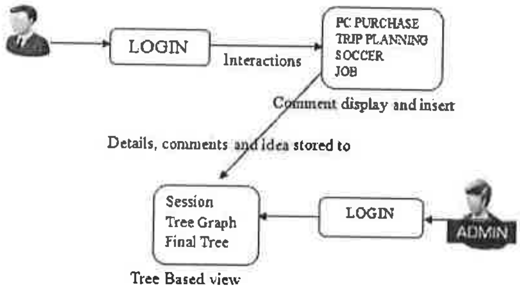


Figure 2 shows the flow diagram of Mining Human Interaction.

3.1. Stochastic Techniques: The visualization systems aim at detecting and visualizing human interactions in meetings, while our work focuses on discovering higher level knowledge about human interaction. There have been several works done in discovering human behavior patterns by using stochastic techniques. A Stochastic Techniques is a collection of random variables; this is often used to represent the evolution of some random value, or system, over time. This is the probabilistic counterpart to a deterministic process. Instead of describing a process which can only evolve in one way, in a stochastic or random process there is some indeterminacy: even if the initial condition is known, there are several directions in which the process may evolve.

3.2. Algorithms for Pattern Discovery: With the representation model and annotated interaction Flows, we generate a tree for each interaction flow and thus build a tree data set. For the purpose of pattern discovery, we first provide the definitions of a pattern and support for determining patterns. In developing our frequent sub tree discovery algorithm, we decided to follow the structure of the algorithm for pattern discovery used for finding frequent item sets, because it achieves the most effective pruning compared with other algorithms.

3.3. Tree pattern mining algorithms: To analyze tree structures and extract interaction flow patterns. An interaction flow that appears frequently reveals relationships between different types of interactions. Mining human interactions is important for accessing and understanding meeting content. A tree-based mining. method is used for discovering frequent patterns of human interaction in meeting discussions. The mining results would be useful for summarization, indexing, and comparison of meeting records. They also can be used for interpretation of human interaction in meetings.

3.4. Interaction Flow Construction: Interaction flow construction create an environment based on the interaction defined and recognized, we now describe the notion of interaction flow and its construction. An interaction flow is a list of all interactions in a discussion session with triggering relationship between them. We create an application based on it. In the application we have authentication process. For authentication process we build Login process, which is used for enter the process and register the new users.

This process is produced for both users and admin process. All users details can be stored in the database elements. So, unwanted users cannot easily access this Login process. Homepage is used for the login. Registration process requires the Name, Details, address, phone number and email id.

3.5. Expressing Opinion: Comments display process is the process of displaying the comments in the user and admin view. But users view is different from the admin view. In user view, the user can view the comments and also enter the comment elements. This comment display is classified by four meeting display in our project are

These processes get the details about the users and get the idea for the topics and also negative and positive comments. Users can view the positive and negative comments of other users are used. So currents users can get the knowledge about that particular topic and also they correct the doubts in their topics.

3.6. Admin Analysis: Admin process is the process that maintains the process and users details. The Main details of the users can not viewed by users, that type of process is maintained by the admin process. Admin process can view the process as tree based structure. So the admin can easily identified by the human interactions. Human interaction process can be viewed by admin by the following structure based elements are various elements which are described in the following modules.

Session tree process is the process that is used to avoid the repeated data in session database. And the process provides the tree based structure. So the admin identified the main problem in the particular topic. All process such as PC purchase, Trip planning, Soccer and job can be viewed by the admin process.

Graph is another process for the admin view. This process is also related to the session tree concept. But is process only provides the separate graph view. So the admin can easily maintain the process.

Final Tree is the process that is fully related to the session tree and graph. This process provides the full view of the user interaction. So details of all users can be easily identified by the admin process.

4. Problem Definition

Discovering a pattern semantic knowledge is significant for understanding and interpreting how the user interact in a meetings like business, commercial & Academic purpose. Common way of capturing the information is through note taking however manually written down the content of a meeting is a difficult task and can result in an ability to both take note and participate in the meeting. Tree based mining method for discovering frequent patterns of human interactions in meeting discussion mining technique analyze the comparison of meeting records and used to understand about human interaction in meetings.

4.1. Tree Based Mining Technique: Finding frequent item sets in data warehouse operation of association rule mining, frequent patterns have many useful applications in markup language marketing banking networking routing. A mining method to extract frequent items of human interaction based mining on the extracted content of interactions. Human interactions such as proposal or giving comments opinions are constructed as a priority like a tree, tree based interaction mining algorithm are designed to analyze the structures of the trees and to extract frequent patterns in a tree dataset. Capturing all of informal meeting information is omitted by using tree based mining approach to extract frequent patterns of human interactions based on the captured content is of human participated meetings, mining results can be used for context purpose meeting semantics also existing meeting capture systems use this technique as a smarter indexing tool to search and access particular semantics of the meetings.

4.2. Human Interaction: Human interaction varies depending on the usage of the meetings or the types of the meetings and task oriented interactions communicative actions that concern the meeting and the group itself. Set of interaction types based on a standard unit scheme comment acknowledge request information ask opinion positive opinion and negative opinion. User proposes an idea with respect to a subject or proposal. Representation use labels for human interactions are abbreviated names of interactions that commenting request information ask opinion giving positive giving negative opinion.

5. MINING ALGORITHMS USED

Supervised learning is to use the available data to build one particular variable of interest in terms of rest of data.

A number of classification algorithms can be used for anomaly detection. proposes the use of ID3 Decision tree classifiers to learn a model that distinguishes the behavior of intruder from the normal users behavior.

Unsupervised learning is where no variable is declared as target the goal is to establish some relationship among all variables

Unsupervised learning studies how systems can learn to represent particular input patterns in a way that reflects the statistical structure of the overall collection of input patterns.

The unsupervised learner brings to bear prior biases as to what aspects of the structure of the input should be captured in the output. In this paper combination of Applications Supervised& Unsupervised has been combined together used to solve the problem of Network Anomaly Data.

A Very rare case both the Techniques have been combined Supervised (Classification): Decision Tree, Bayesian classification, Bayesian belief networks, neural networks etc. are used in data mining based applications.

A. Classification Techniques:

In Classification, training examples are used to learn a model that can classify the data samples into known classes. The Classification process involves following steps:

a. Create training data set.
b. Identify class attribute and classes.
c. Identify useful attributes for classification (relevance analysis).
d. Learn a model using training examples in training set.
e. Use the model to classify the unknown data samples.

Unsupervised (Clustering): Association Rules, Pattern Recognition, Clustering Technique

The paper clustering Technique is one of the media to Network Anomaly data

B. Clustering Technique

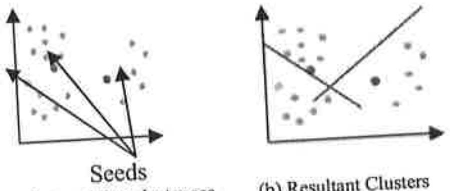
Cluster is a number of similar objects grouped together. It can also be defined as the organization of dataset into homogeneous and/or well separated groups with respect to distance or equivalently similarity measure. Cluster is an aggregation of points in test space such that the distance between any two points in cluster is less than the distance between any two points in the cluster and any point not in it. There are two types of attributes associated with clustering, numerical and categorical attributes. Numerical attributes are associated with ordered values such as height of a person and speed of a train. Categorical attributes are those with unordered values such as kind of a drink and brand of car.

Clustering is available in flavors of

- Hierarchical
- Partition

In hierarchical clustering the data are not partitioned into a particular cluster in a single step. Instead, a series of partitions takes place, which may run from a single cluster containing all objects to n clusters each containing a single object. Hierarchical Clustering is subdivided into agglomerative methods, which proceed by series of fusions of the n objects into groups, and *divisive* methods, which separate n objects successively into finer groupings

For the partitional can be of K-means & K-mediod. The purpose solution is based on K-means clustering combine with Id3 Decision Tree type of Classification under mentioned section describes in details of K-means & Decision Tree. K-means is a centroid based technique Each cluster is represented by the center of gravity of the cluster so that the intra cluster similarity is high and inter cluster similarity is low. This technique is scalable and efficient in processing large data sets because the computational complexity is O(nkt) where n-total number of objects, k is number of clusters, t is number of iterations and k<<n and t<<n.



(a) Un clustered Data Instances (b) Resultant Clusters

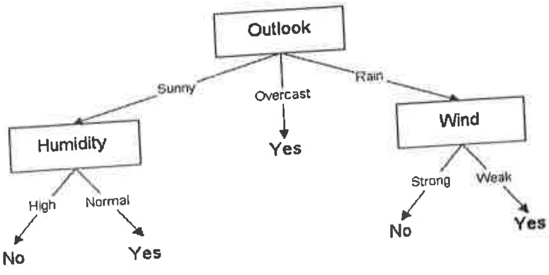
Figure 1 : Formation of clusters using seed points

C. K-mean algorithm:

1. Select k centroids arbitrarily (called as seed as shown in the figure) for each cluster C_i, i ε [1, k]
2. Assign each data point to the cluster whose centroid is closest to the data point.
3. Calculate the centroid C_i of cluster C_i, i ε [1, k] In short
4. Repeat steps 2 and 3 until no points change between clusters. A major disadvantage of K means is that one must specify the clusters in advance and further the algorithm is very sensitive of noise, mixed pixels and outliers. The definition of means limit the application to only numerical variables. We choose k-means because it is data driven with relatively few assumptions on the distributions of underlying dat.

D. Decision Tree

Decision tree support tool that uses tree-like graph or models of decisions and their consequences, including event outcomes, resource costs, and utility. Commonly used in operations research, in decision analysis help to identify a strategy most likely to reach a goal. In data mining and machine learning, decision tree is a predictive model, that is mapping from observations about an item to conclusions about its target value. The machine learning technique for inducing a decision tree from data is called decision tree learning.



The example of fig(2) is taken from[11]

In above fig(2) tree is classified into five leaf nodes. In a decision tree, each leaf node represents a rule. The following rules are as follows in figure(2) Rule **1:** If it is sunny and the humidity is high then do not play. Rule2 : If it is sunny and the humidity is normal then play. Rule3 : If it is overcast, then play. Rule4 : If it is rainy and wind is strong then do not play. Rule5 : If it is rainy and wind is weak then play.

E. ID3 Decision Tree

Iterative Dichotomiser is an algorithm to generate a decision tree invented by Ross Quinlan, based on Occam's razor. It prefers smaller decision trees(simpler theories) over larger ones. However it does not always produce smallest tree, and therefore heuristic. The decision tree is used by the concept of Information Entropy

The ID3 Algorithm is:

1) Take all unused attributes and count their entropy concerning test samples

2) Choose attribute for which entropy is maximum

3) Make node containing that attribute

```
ID3 (Examples, Target _ Attribute, Attributes)
•  Create a root node for the tree
•  If all examples are positive, Return the single-node tree
   Root, with label = +.
•  If all examples are negative, Return the single-node tree
   Root, with label = -.
•  If number of predicting attributes is empty, then Return
   the single node tree Root, with label = most common
   value of the target attribute in the examples.
•  Otherwise Begin
o  A = The Attribute that best classifies examples.
o  Decision Tree attribute for Root = A.
o  For each possible value, $v_i$, of A,
▪        Add a new tree branch below Root,
         corresponding to the test A = $v_i$.
▪        Let Examples($v_i$), be the subset of
         examples that have the value $v_i$ for A
▪        If Examples($v_i$) is empty
         common target value in the examples
▪        Else below this new branch
         add the sub tree ID3 (Examples($v_i$), Target_ Attribute,
         Attributes – {A}
•  End
•  Return Root
```

F. Association Mining: Association mining is patterns in data invented by databases or is the business field where discovering of purchase patterns or association between products is useful for decision making and effective marketing. Finding all relevant occurrence relationship is called association, it is market basket data analysis to discover items purchased by customers in a market are associated.

Association mining algorithm developed with different mining efficiencies, which find the same set of rules though their computational efficiencies and memory requirements may be different in two steps. Frequent item is an item set that has transaction support, confident association rule is with confidence.

Market basket data is different data can be tailored to fit to the definition of transactional databases so that association rule mining algorithm can be applied to them. Text document can be seen as transaction data. Each document is a transaction and each distinctive to convert a table data to transaction data if each attribute in table takes categorical values.

Association mining is the threshold used to prune the search space and to limit the number of frequent item set and rules generated. But using only a single implicitly assumes that all items in the data are of the same nature similar frequencies in the database. In some other applications items appear very frequently in the data that perform if the minsup is set too high not find rules that involve infrequent items or items are rare in the data, to find rules that involve both frequent and rare items have to set the minsup very low.

Lk: set of frequent item set of size k with min support

Ck: set of candidate item set of size k potentially frequent itemset

L1 is frequent items for (k=1; Lk! =Ø; K++) do begin s

6. COMPARATIVE STUDY

Users meeting capture systems could use this technique as a smarter indexing tool to search and access only particular semantics of the meetings. This work focuses on only lower level knowledge about human interaction. The process didn't have any key features. So it not compares two meeting discussions. The process only gets the positive and negative comments from the users. So further process to be discussed only by the admin. So complex of the topic should not be identified easily. Sometimes this process not provides the semantic information and produces redundant data which is complex to handle, Identification of negative points in topic is very tough and increases the repeated data. Compare early system our work a mining method to extract frequent patterns of human interaction based on the captured content of face-to-face meetings. The work focuses on discovering higher level knowledge about human interaction. In our proposed system T-pattern technique is used to discover hidden time patterns in human behavior. We conduct analysis on human interaction in meetings and address the problem of discovering interaction patterns from the perspective of data mining. It extracts simultaneously occurring patterns of primitive actions such as gaze and speech. We discover patterns of interaction flow from the perspective of tree-based mining rather than using simple statistics of frequency. The main features of the process are user can also provide the idea about the topic. So admin can easily solve the problem based on users needed extracts data simultaneously & problems occurred in the process is easily solved by the admin.

7. CONCLUSION

Mining tree method for discovering frequent patterns of human interaction in communication of business requirement, analysis shows the comparison of existing system mining algorithms used to extract meeting records. It is explore tree mining for hidden interaction pattern discovery. Communication is current meetings are task oriented is to discover frequent interaction trees and the behavior of the algorithms on the data set using the behavior. Several applications based on the discovered patterns which allow is used to extract the temporal pattern in case of the time period along with the tree mining and finally plan to extract more meeting content in large volume.

REFERENCES

[1]. Christian Becker, Zhiwen Yu,"Tree-Based mining for Discovering Patterns of Human Interaction in Meetings", Proc. Eighth IEEE Int'l Conf. Knowledge Discovery and Data Mining (PAKDD'10), pp. 107-115,2012.

[2]. Z.Yu, Y. Nakamura,"Smart Meeting Systems: A Survey of State-of-the-Art and Open Issues", ACM Computing Surveys, Vol. 42, No. 2, article 8, Feb. 2010

[3]. J. Han, M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, March 2006.

[4]. A. K. Jain and R. C. Dubes. Algorithms for Clustering Dasta. Printice Hall, 1988.

[5]. A. K. Jain, M. Narasimha Murty, and P.J. Flynn. Data Clustering: A Review. ACM Computing Surveys, 31(3):264–323, 1999.

[6]. V.V. Srinivas and N. Ramasubramanian, (2011), "Understanding the performance of multi-core

[7]. platforms", International Conference on Communications, Network and Computing, CCIS-142,pp. 22-26.

About the authors:

RameshBabu Varugu B.Tech Computer Science Engineering from Lakkireddy Balreddy Engineering College M.Tech Information Technology from Gurunank Engineering College. Having eight years of experience in Academic currently working as Asst Prof in Annamacharya Institute of science & Technology. He has guided many UG & PG students interested subjects include Network Security, Data Mining, Cloud Computing.

C Shanker B.Tech Computer Science Engineering from Vijay Rural Engineering College M.Tech Computer Science from J.B.Institute of Engineering & Technology. Currently he working as Asst Prof in Sri Indu College of Engineering & Technology having six years of Academic experience guided many UG & PG students. His interested subjects include Data Mining Software Engineering Network Security & Cloud Computing.

Design & Development of an Effective Video Streaming Framework using Cloud Computing Technology

K. Narsimhulu[1], K.V.S. Sudhakar[2], N. Yadagiri[3], Prof.Dr.G.Manoj Someswar[4]
1. Suprabhath College of Engineering & Technology, Sheriguda, Ibrahimpatnam, RR District., Telangana State, India.
2. Associate Professor, Department of CSE, Suprabhath College of Engineering & Technology, Sheriguda, Ibrahimpatnam, RR District., Telangana State, India.
3. HOD, Dept. of CSE, Suprabhath College of Engineering & Technology, Sheriguda, Ibrahimpatnam, RR District., Telangana State, India.
4. Principal & Professor, Department of CSE, Anwar-ul-uloom College of Engineering & Technology, Yennepally, Vikarabad, Telangana State, India

Abstract: While demands on video traffic over mobile networks have been souring, the wireless link capacity cannot keep up with the traffic demand. The gap between the traffic demand and the link capacity, along with time-varying link conditions, results in poor service quality of video streaming over mobile networks such as long buffering time and intermittent disruptions. Leveraging the cloud computing technology, we propose a new mobile video streaming framework, dubbed AMES-Cloud, which has two main parts: AMoV (adaptive mobile video streaming) and ESoV (efficient social video sharing). AMoV and ESoV construct a private agent to provide video streaming services efficiently for each mobile user. For a given user, AMoV lets her private agent adaptively adjust her streaming flow with a scalable video coding technique based on the feedback of link quality. Likewise, ESoV monitors the social network interactions among mobile users, and their private agents try to prefetch video content in advance. We implement a prototype of the AMES-Cloud framework to demonstrate its performance. It is shown that the private agents in the clouds can effectively provide the adaptive streaming, and perform video sharing (i.e., prefetching) based on the social network analysis.

Keywords: Efficient Social Video Sharing, Adaptive Mobile Video Streaming, Wireless Link Bandwith, Scalability, Adaptability, videos in the clouds

I. INTRODUCTION

Over the past decade, increasingly more traffic is accounted by video streaming and downloading. In particular, video streaming services over mobile networks have become prevalent over the past few years. While the video streaming is not so challenging in wired networks, mobile networks have been suffering from video traffic transmissions over scarce bandwidth of wireless links. Despite network operators' desperate efforts to enhance the wireless link bandwidth (e.g., 3G and LTE), soaring video traffic demands from mobile users are rapidly overwhelming the wireless link capacity. While receiving video streaming traffic via 3G/4G mobile networks, mobile users often suffer from long buffering time and intermittent disruptions due to the limited bandwidth and link condition fluctuation caused by multi-path fading and user mobility.[1] It is crucial to improve the service quality of mobile video streaming while using the networking and computing resources efficiently.To improve the service quality of mobile video streaming on two aspects:

Scalability: Mobile video streaming services should support a wide spectrum of mobile devices; they have different video resolutions, different computing powers, different wireless links (like 3G and LTE) and so on. [2]

Adaptability: To address this issue, we have to adjust the video bit rate adapting to the currently time-varying available link bandwidth of each mobile user. Such adaptive streaming techniques can effectively reduce packet losses and bandwidth waste.

```
i+1 (or BWestimate
i+1 = BWpractical
i )
k=0
BWEL=0
repeat
k++
if k >= j break
BWEL=BWEL + RELk
until BWEL >= BWestimate
i+1
□RBL
Transmit BLi+1 and EL1
i+1, EL2
i+1,..., ELk□
i+1
Monitor BWpractical
i+1
i++
until All video segments are transmitted
```

II. ARCHITECTURE

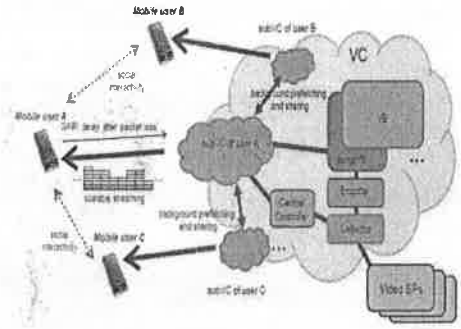
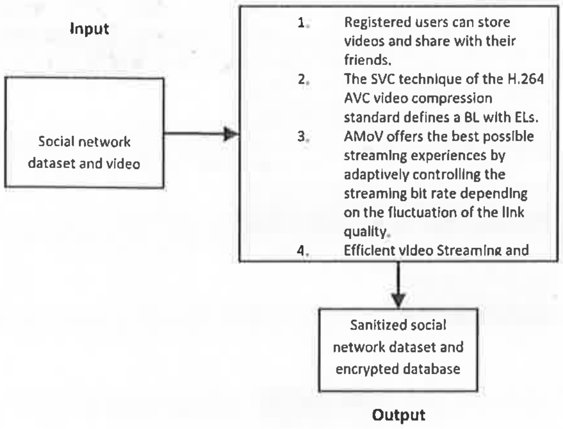


Figure 1: Sharing of Videos and Messages



Input

Social network dataset and video

1. Registered users can store videos and share with their friends.
2. The SVC technique of the H.264 AVC video compression standard defines a BL with ELs.
3. AMoV offers the best possible streaming experiences by adaptively controlling the streaming bit rate depending on the fluctuation of the link quality.
4. Efficient video Streaming and

Sanitized social network dataset and encrypted database

Output

Figure 2: Functional Diagram

III. EXISTING SYSTEM

Cloud computing promises lower costs, rapid scaling, easier maintenance, and service availability anywhere, anytime, a key challenge is how to ensure and build confidence that the cloud can handle user data securely. A recent Microsoft survey found that "58 percent of the public and 86 percent of business leaders are excited about the possibilities of cloud computing. But more than 90 percent of them are worried about security, availability, and privacy of their data as it rests in the cloud.[3]

Proposed System

We propose an adaptive mobile video streaming and sharing framework, called AMES-Cloud, which efficiently stores videos in the clouds (VC), and utilizes cloud computing to construct private agent (subVC) for each mobile user to try to offer "non-terminating" video streaming adapting to the fluctuation of link quality based on the Scalable Video Coding technique. Also AMES-Cloud can further seek to provide "nonbuffering"experience of video streaming by background pushing functions among the VB, subVBs and localVB of mobile users. We evaluated the AMES-Cloud by prototype implementation and shows that the cloud computing technique brings significant improvement on the adaptivity of the mobile streaming. We ignored the cost of encoding workload in the cloud while implementing the prototype.

IV. INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

□ What data should be given as input?

□ How the data should be arranged or coded?

□ The dialog to guide the operating personnel in providing input.

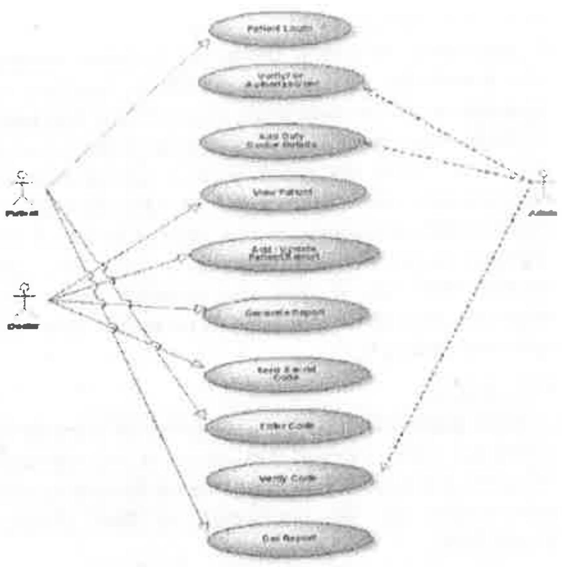


Figure 5: Use Case Diagram

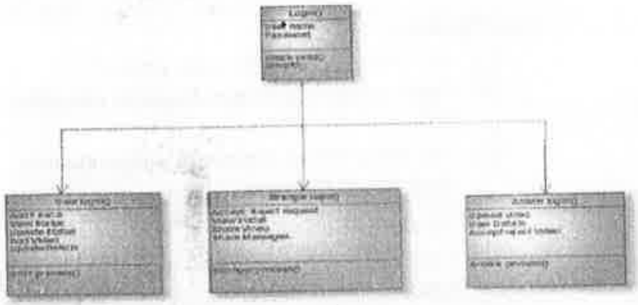


Figure 6: Class Diagram

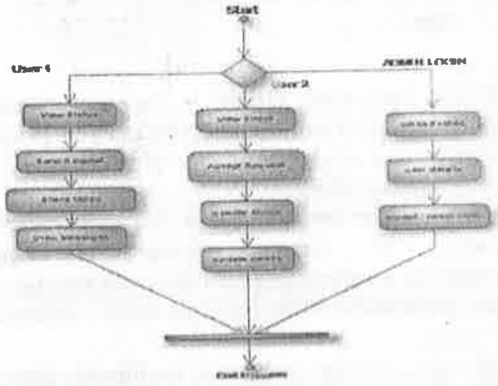


Figure 7: Activity Diagram

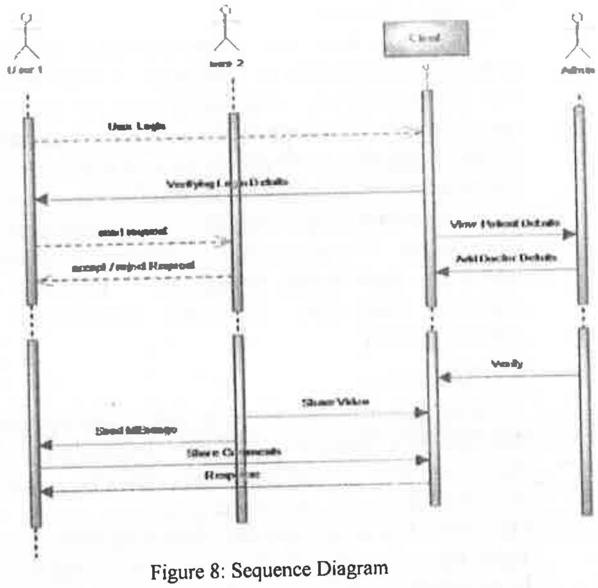


Figure 8: Sequence Diagram

IX. SYSTEM TESTING

The purpose of testing is ,to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

X. TYPES OF TESTS

Unit Testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.[4]

end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

XI. CONCLUSION

In this research paper, we discussed our proposal of an adaptive mobile video streaming and sharing framework, called AMES-Cloud, which efficiently stores videos in the clouds (VC), and utilizes cloud computing to construct private agent (subVC) for each mobile user to try to offer "non-terminating" video streaming adapting to the fluctuation of link quality based on the Scalable Video Coding technique. Also AMES-Cloud can further seek to provide "nonbuffering" experience of video streaming by background pushing functions among the VB, subVBs and localVB of mobile users. We evaluated the AMES-Cloud by prototype implementation and shows that the cloud computing technique brings significant improvement on the adaptivity of the mobile streaming. The focus of this paper is to verify how cloud computing can improve the transmission adaptability and prefetching for mobile users. We ignored the cost of encoding workload in the cloud while implementing the prototype. As one important future work, we will carry out large-scale implementation and with serious consideration on energy and price cost. In the future, we will also try to improve the SNS-based prefetching, and security issues in the AMES-Cloud.

Future Enhancement

As one important future work, we will carry out large-scale implementation and with serious consideration on energy and price cost. In the future, we will also try to improve the SNS-based pre fetching, and security issues in the AMES-Cloud.

SNS-Based Pre Fetching

In order to improve quality of service (QoS) of wireless environment in scene-based mobile social network service (mobile SNS), a novel pre fetching. Utilizing this method, pre fetching decisions are made based on the relationship among users and access histories recorded in the client. When the user is viewing the scene or other components, client is allowed to pre fetch scenes from the server that the user will potentially access soon. Experiment has been done on a prototype called scene life system developed by us. The results show that the proposed method can significantly reduce scene access time, and then provide better QoS.

REFERENCES

[1] CISCO, "Cisco Visual Networking Index : Global Mobile Data Traffic Forecast Update , 2011-2016," Tech. Rep., 2012.

[2] Y. Li, Y. Zhang, and R. Yuan, "Measurement and Analysis of a Large Scale Commercial Mobile Internet TV System," in *ACM IMC*, pp. 209–224, 2011.

[3] T. Taleb and K. Hashimoto, "MS2: A Novel Multi-Source Mobile-Streaming Architecture," in *IEEE Transaction on Broadcasting*, vol. 57, no. 3, pp. 662–673, 2011.

[4] X. Wang, S. Kim, T. Kwon, H. Kim, Y. Choi, "Unveiling the BitTorrent Performance in Mobile WiMAX Networks," in *Passive and Active Measurement Conference*, 2011.

[5] A. Nafaa, T. Taleb, and L. Murphy, "Forward Error Correction Adaptation Strategies for Media Streaming over Wireless Networks," in *IEEE Communications Magazine*, vol. 46, no. 1, pp. 72–79, 2008.

[6] J. Fernandez, T. Taleb, M. Guizani, and N. Kato, "Bandwidth Aggregation-aware Dynamic QoS Negotiation for Real-Time Video Applications in Next-Generation Wireless Networks," in *IEEE Transaction on Multimedia*, vol. 11, no. 6, pp. 1082–1093, 2009.

[7] T. Taleb, K. Kashibuchi, A. Leonardi, S. Palazzo, K. Hashimoto, N. Kato, and Y. Nemoto, "A Cross-layer Approach for An Efficient Delivery of TCP/RTP-based Multimedia Applications in Heterogeneous Wireless Networks," in *IEEE Transaction on Vehicular Technology*, vol. 57, no. 6, pp. 3801–3814, 2008.

[8] K. Zhang, J. Kong, M. Qiu, and G.L Song, "Multimedia Layout Adaptation Through Grammatical Specifications," in *ACM/Springer Multimedia Systems*, vol. 10, no. 3, pp.245–260, 2005.

[9] M. Wien, R. Cazoulat, A. Graffunder, A. Hutter, and P. Amon, "Real-Time System for Adaptive Video Streaming Based on SVC," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1227–1237, Sep. 2007.

[10] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

IJETCR
Open Access

ISSN: 2348 - 2117

International Journal of Engineering Technology and Computer Research (IJETCR)
Available Online at www.ijetcr.org
Volume 2; Issue 6; Page No. 55-59

Analysis of Cloud Data Mining & Emergence of Self-Destructing Technique to Archive Data

Varugu Ramesh Babu, V.Rama Krishna

Associate Professor, Ananamacharya Institute of science & Technology.

Associate Professor, Ananamacharya Institute of science & Technology.

ABSTRACT

Cloud database Storage is a model of networked enterprise storage where data is stored in virtualized pools of storage which are generally hosted by vendors it may contains large data centers. The information like education notes and business deals documents that could be misused by hacker's unauthorized users. These data copied and archived by cloud storage service often with users and control using self-destructing. Self-destructing protecting the user data in secure way all the data and their copies become destructed or unreadable after a user specified time without any user intervention. This works presents the self-destructing survey interact with cryptography technique with active storage meets the privacy concerns. The functionality compares the earlier self-destructing with our proposed analysis shamir secret sharing algorithmis efficient not used in any earlier data mechanism.

Keywords: Destructing, Cloud Service Provider, Storage device, Cloud databases.

INTRODUCTION:

Database management systems are an integral and indispensable component in most computing organizations today with the advent of hosted cloud computing and storage. Cloud computing is the evolution of internet based computing provided a common infrastructure for applications static web pages began to add interactivity hosted applications like Hotmail more user configuration renamed software as a service. With a growing number of companies looking to get on the software as a service opportunity Amazon released web services that enable companies to operate their own software as a service applications. Cloud database usage patterns are evolving and business adoption of these technologies accelerates that evolution cloud databases serviced consumer applications these early applications put a priority on read access because the ratio of reads to writes was very high. Consumer centric cloud database applications have been evolving with the adoption of web 2.0 technologies user generated content particularly in the form of social networking for example consumer centric cosmetics website if the user does a search for a certain shade of makeup powder it is important that the results be delivered instantaneously to keep the user engaged so she does not click on another cosmetics site. If the site said that the chosen makeup powder is in inventory and completed the sale it would not be the end of the world to later find out that a result of inconsistent data that makeup powder was not really in inventory.Cloud database is a database that consists of cloud computing like Google Microsoft Salesforce Rackspace Amazon etc, cloud database management system are designed to satisfy applications such as availability of a service Data confidentiality flexible query interface. Cloud architecture consists of layers Manageability layer deals with managing various users keeps the record of the time a particular user uses the cloud database.

Security layer provides user authentication mechanism with the help of users' id and passwords should be accepted as being legal one.

Transparency provides transparences to the users of cloud database where it means that the physical of data is not known to the users various types of real time applications easier.

Conceptual is heterogeneity among different databases like SQL DB2 Oracle a logical structure of the entire database deals with the internal processing on data as cloud deals with various types of data here users need to combine the traditional data with the data that are placed on the cloud so various types of systems are required for cloud database that provides all functions.

Interoperability means operate irrespective of their underlying databases for example if customer A wants to share data with another customer with B they are able to share the data irrespective of their underlying different databases of different vendors with the help of this layer.

Cloud computing as a utility that has recently attracted significant features people used terminals to connect to powerful mainframes shared by many users, the standalone personal computers became powerful enough to satisfy users daily work and computer networks allowed multiple computers to connect to each other, the cloud computing allows the exploitation of all available resources on the internet in a scalable way.

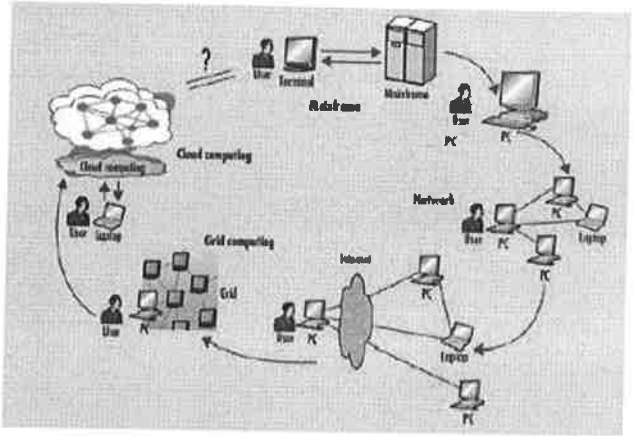


Figure 1: shows the Cloud Storage Environment

Cloud computing is amodel for enabling ubiquitous convenient on demand network access to shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or serive provider interaction. Data mining represents finding useful patterns or trends through large amounts of data defined as atype of database useful patterns or relationships in a group of data uses advanced statistical methods such as cluster analysis and sometimes employs artificial intelligence or neural networks. Data miningthe extraction of hidden predictive information from large databases is a powerful with great potential to help companies focus on the most important information in their datawarehouse. Cloud computing denotes the new trend in internet services that rely on clouds of servers to handledtasks.

SECTION II

Related Work: When use the system as detection might become complex study of new idea for sharing and protecting privacy system a secret key is divided and stored in a point to point system with distributed hash tables. With joining and exiting of the point to point node the system can maintain secret keys according to characteristics of point to point the distributed has table will refresh every node after every certain hours with secret sharing algorithm when will not get enough parts of a key the person will no decrypt data encrypted with this key means that key is destroyed and the data cannot be recovered some attacks to characteristics of point topoint are challenges of system uncontrolled in how long the key can survive. System used for creating message that automatically self-destruct after certain period of time which integrates cryptographic techniques with global scale point to point distributed hash tables. Distributed hash table have the property to discard data older than a certain age. In this the key permanently lost and the encrypted data is system each message is encrypted with a random key and storing share of the key in a large public distributed hash tables.

Self-destructive system defines two new modules associated with each secret key part and each secret key part has its own survival time parameter. In the self-destructive system can meet the requirements of self-destructing data with controllable survival time while users can use this system as a general object storage system. Apply some load balancing and round trip algorithms an active storage object derives from a user object and has a time to live value property which used to trigger the self-destruct operation. The time to live value of a user object has the property infinite so the user object will not be deleted until a user deletes it manually on the other hand the time to live value of an active storage object is limited so an active object will be deleted when the value of the associated policy object is true. Secure delete sensitive data and reduce the negative impact of performance due to deleting operation the required secure deletion of all the files is not great so if these parts of the file update operation changes.Self-destruction data is implemented by encrypting data with a key and that information is needed to reconstruct the decryption key with many parties local data destruction approach will not work in the cloud storage because the number of backups or archives of the data that is stored in the cloud is unknown and some nodes preserving the backup data have been offline. System creating messages that automatically self-destruct after a period of time it gets integrates cryptographic techniques with global scale peer-to-peer distribution has table.

SECTION III

3.Number of network intrusions have been found till now, each of which utilizes one or more security vulnerabilities in TCP/IP protocol specifications. These intrusions include IP source address spoofing, TCP sequence number prediction as mentioned earlier and other intrusions like SYN flooding, DNS misuse, Ping of Death, or some Java-related attacks. However, based on the intrusion patterns and impacts to the victim systems, intrusions into two main categories: denial of service and spoofing.The lifeblood of today's world is information. The denial-of-service intrusions attempt to prevent or delay access to the information or the information processing systems. The basic idea behind this type of intrusion is to tie up a service provider with bogus requests in order to render it unreliable or unusable.

Network –based intrusion detection system[NIDS]][6] that tries to detect malicious activity such as denial of service attacks, port scan or even attempts to crack into computer by monitoring network traffic. NIDS does this by reading all incoming packets and trying to find number of TCP connection requests to a very large number of different ports is observed, one could assume that there is someone conducting a port scan of some or all of the computers in the network. It mostly tries to detect incoming shell codes in the same manner that an ordinary intrusion detection system does. Often inspecting valuable information about an ongoing intrusion can be learned from outgoing or local traffic and also work with other systems as well, for example update some firewalls blacklist with the IP address of computers used by suspected crackers.

Host-based intrusion detection system [HIDS] [4] monitors parts of the dynamic behavior and the state of computer system, dynamically inspects the network packets. A HIDS could also check that appropriate regions of memory have not been modified, for example- the system-call table comes to mind for Linux and various v table structures in Microsoft windows. For each object in question usually remember its attributes (permissions, size, modifications dates) and create a checksum of some kind (an MD5, SHA1 hash or similar) for the contents, if any, this information gets stored in a secure database for later comparison (checksum-database). At installation time- whenever any of the monitored objects change legitimately- a HIDS must initialize its checksum-database by scanning the relevant objects. Persons in charge of computer security need to control this process tightly in order to prevent intruders making un-authorized changes to the database.

Protocol-based intrusion detection system [PIDS][4] typically installed on a web server, monitor the dynamic behavior and state of the protocol, typically consists of system or agent that would sit at the front end of a server, monitoring the HTTP protocol stream. Because it understands the HTTP protocol relative to the web server/system it is trying to protect it can offer greater protection than less in-depth techniques such as filtering by IP address or port number alone, however this greater protection comes at the cost of increased computing on the web server and analyzing the communication between a connected device and the system it is protecting.

Application protocol based intrusion detection system [APIDS][4] will monitor the dynamic behavior and state of the protocol and typically consists of a system

or agent that would sit between a process, or group of servers, monitoring and analyzing the application protocol between two connected devices.

SECTION IV

4. Problem Definition: Computer networkis a system that we can perform development, software applications, used to transfer data packets,unauthorized attacks intrusion detection became the anomaly, installation of antivirus protection software the system became more complex, to avoid all these detection and protection over the computer applications data packets will not provide secure for our data. Cloud storage service is technique that provides destructing when the application of development is to be done.Use of destructing is information available in the cloud could not miscreant or court law these data is cached copied and archived by cloud service providers often without user's authorization and control. If we retrieve at any time all the data and their copies become destructed or unreadable after a user specified time without any user intervention.

Computer Operating system kernel code as code that to be executing a service method should be implemented in user space with libraries functions is used by code in user specific functions. Tools to develop software system in user space much safe to debug code in user spam than in kernel space. The method process takes long time a complicated task so implementing code of a method in user space have advantage of performance of the system. The system might crash with an error in kernel code but if error occurs in code of user space self-destruct method object is a method with arguments which specifies the device object to be destructed.

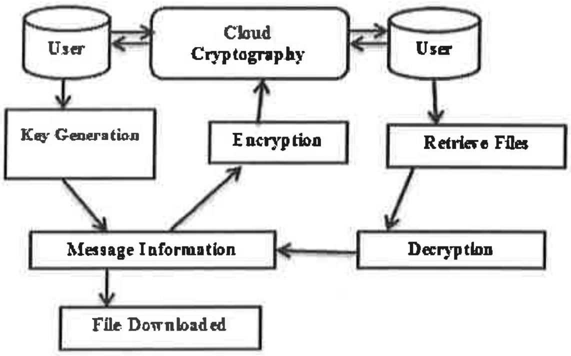


Figure :2 De-structing cloud Storage

SECTION V

5.1. Shamir algorithm: Secret sharing is a method in cryptography for distributing a secret among a group of participants each of which is allocated to share a secret. Secret can only be reconstructed when the shares are combined together or individual share are of use on their own. Sharing a secret gives control and removes single point vulnerability, independent share holder cannot change or access the data.

A goal is divide some of data for example D into n pieces D1, D2....Dn and a knowledge of any k or more D pieces makes D easily computable, knowledge of any k-1 or fewer pieces leaves D completely undetermined is called (k,n) threshold scheme if k=n then all participants are required together reconstruct the secret.

Suppose we want to use (k,n) threshold scheme to share our secret S where k<n at random (k-1) coefficients a1,a2,a3...ak-1 and let S be the a0.

$$f(x) = a_0 + a_1 x + a_2 x^2 + ... + a_{k-1}{}^{k-1}$$

Construct n points (I,f(i)) where i=1,2,..n

Given any subset of k of these pairs we can find the coefficients of the polynomial by interpolation and then evaluate a0=S, which is the secret.

5.2. Cloud Data Sharing &Storage Analysis: users, who have data to be stored in the cloud and rely on the cloud for data computation, consist of both individual consumers and organizations.

Cloud Service Provider (CSP): a CSP, who has significant resources and expertise in building and managing distributed cloud storage servers, owns and operates live Cloud Computing systems.

Third Party Auditor (TPA): an optional TPA, who has expertise and capabilities that users may not have, is trusted to assess and expose risk of cloud storage services on behalf of the users upon request.

5.2.1. File Retrieval and Error Recovery

Since our layout of file matrix is systematic, the user can reconstruct the original file by downloading the data vectors from the first m servers, assuming that they return the correct response values. Notice that our verification scheme is based on random spot-checking, so the storage correctness assurance is a probabilistic one. We can guarantee the successful file retrieval with high probability. On the other hand, whenever the data corruption is detected, the comparison of pre-computed tokens and received response values can guarantee the identification of misbehaving server(s).

5.2.2. Third Party Auditing

In case the user does not have the time, feasibility or resources to perform the storage correctness verification, he can optionally delegate this task to an independent third party auditor, making the cloud storage publicly verifiable. However, as pointed out by the recent work, to securely introduce an effective TPA, the auditing process should bring in no new vulnerabilities towards user data privacy. Namely, TPA should not learn user's data content through the delegated data auditing.

5.2.3. Cloud Operations

Update Operation

In cloud data storage, sometimes the user may need to modify some data block(s) stored in the cloud, we refer this operation as data update. In other words, for all the unused tokens, the user needs to exclude every occurrence of the old data block and replace it with the new one.

Delete Operation

Sometimes, after being stored in the cloud, certain data blocks may need to be deleted. The delete operation we are considering is a general one, in which user replaces the data block with zero or some special reserved data symbol. From this point of view, the delete operation is actually a special case of the data update operation, where the original data blocks can be replaced with zeros or some predetermined special blocks.

Append Operation

In some cases, the user may want to increase the size of his stored data by adding blocks at the end of the data file, which we refer as data append. We anticipate that the most frequent append operation in cloud data storage is bulk append, in which the user needs to upload a large number of blocks (not a single block) at one time.

SECTION V

6. Comparative Study: Cloud storage services for a user to store data to avoid problem that can raised by the centralized trusted third party of self-destructing is to protect the user key and provide the functions of self-destructing data. The system contains clients and vendor party data storage and self-destructing. The process to store data has no change, cryptography is applied to upload and download data from cloud storage it mainly runs in kernel mode and it can mount a remote file system to local machine. The input full path of file key file and the life time for key parts system encrypts data and uploads encrypted data system prompts creating active object are successful afterwards and that means the uploading file gets completed. Personal data stored in the Cloud may contain account numbers, passwords, notes, and other important information that could be used and misused by a miscreant, a competitor, or a court of law. These data are cached, copied, and archived by Cloud Service Providers (CSPs), often without users' authorization and control. Self-destructing data mainly aims at protecting the user data's privacy. All the data and their copies become destructed or unreadable after a user-specified time, without any user intervention. Besides, the decryption key is destructed after the user-specified time. These data are cached, copied, and archived by Cloud Service Providers (CSPs), often without users' authorization and control. Self-destructing data mainly aims at protecting the user data's privacy. All the data and their copies become destructed or unreadable after a user-specified time, without any user intervention. Besides, the decryption key is destructed after the user-specified time. we present Self-destructing, a system that meets this challenge through a novel integration of cryptographic techniques with active storage techniques based on T10 OSD standard. We implemented a proof-of-concept Self-destructing prototype. Through functionality and security properties evaluation of the Self-destructing prototype, the results demonstrate that Self-destructing is practical to use and meets all the privacy-preserving goals described above. Compared with the system without self-destructing data mechanism, throughput for uploading and downloading with the proposed Self-destructing acceptably decreases by less than 72%, while latency for upload/download operations with self-destructing data mechanism increases by less than 60%. Compared with the system without self-destructing data mechanism, throughput for uploading and downloading with the proposed Self-destructing acceptably decreases by less than 72%, while latency for upload/download operations with self-destructing data mechanism increases by less than 60%.

CONCLUSION VI

Data security has become increasingly important in the Cloud database. The analysis of paper presents new technique for protecting data privacy from unauthorized users who retroactively obtain, through legal or other means, a user's of cloud storage data and private decryption keys. Comparative study shows the fixed data timeout and large replication factor present challenges for a self-destruction data system.

REFERENCE:

1. Peter Mell, and Timothy Grance, "The NIST Definition of Cloud Computing", The National Institute of Standards and Technology, USA, 2011, Link: http:// csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.
2. IT Strategists, "Top Cloud Computing Companies and Key Features", Link: http://www.itstrategists.com /Top-Cloud-Computing-Companies.aspx.
3. Merriam-Webster Dictionary, "Definition of data mining", Link: http://www.merriam-webster.com/ dictionary/data%20mining.
4. C. Kruegel and G. Vigna, "Anomaly Detection of Web-Based Attacks," Proc. ACM Conf. Computer and Comm. Security, Oct. 2003.

	Ramesh Babu Varugu B.Tech Computer Science Engineering from LakkireddyBalreddy Engineering College M.Tech Information Technology from Gurunank Engineering College. Having eight years of experience in Academic currently working as Asst Prof in Annamacharya Institute of Science & Technology and guided many UG & PG students His research areas include Network Security, Data Mining, Cloud Computing.
	V.Rama Krishna B.Tech Computer Science engineering from Gulberga University M.Tech Computer Science Engineering from VT University Karnataka. Having twelve years of experience in Academic currently working as Assoc Prof in Ananamacharya Institute of science & Technology. He has guided many UG & PG student's interested subjects include Network Security, Data Mining, Cloud Computing.

S.No: ①

MEASUREMENT OF TRAFFIC CONGESTION ON HIGH DENSE URBAN CORRIDORS IN HYDERABAD CITY

CHINNAM TILAK
M Tech Student, Transportation Engineering,
Vishwabharathi College Of Engineering,
Kukatpally, Hyd-500085
E-Mail: Thilakkanna96@Gmail.Com

DR. R. RAMESH REDDY (Professor)
Transportation Engineering Department
E-Mail: rameshlalitha@yahoo.com

ABSTRACT:
Traffic congestion has been one of major issues that most metropolises are facing. It is believed that identification of congestion is the first step for selecting appropriate mitigation measures. Congestion - both in perception and in reality - impacts the movement of people. Traffic congestion wastes time, energy and causes pollution. There are broadly two factors, which effect the congestion; (a) micro-level factors (b) macro-level factors that relate to overall demand for road use. Congestion is 'triggered' at the 'micro' level (e.g. on the road), and 'driven' at the 'macro' level. Micro level factors are, for example, many people want to move at the same time, too many vehicles for limited road space. On the other side, macro level factors are e.g. land-use patterns, car ownership trends, regional economic dynamics, etc. is paper gives an overview and presents the possible ways to identify and measure metrics for urban arterial congestion. A systematic review is carried out, based on measurement metrics such as speed, travel time/delay and volume and level of service. e review covers distinct aspects like definition; measurement criteria followed by different countries/organizations. The strengths and weaknesses of these measures are discussed. Further, a short critique of measurement criteria is presented.

This study aimed to analyze traffic congestion in urban road networks. The speed performance index was adopted to evaluate the existing road network conditions of congestion, then road segment and network congestion indexes were introduced to respectively measure the congestion levels of urban road segment and network. Urban traffic congestion has different typical characteristics under the influence of different conditions, such as different day of week, holiday and weather etc. It is necessary to set up the relationships between traffic congestion patterns and those influencing factors, when we conducting macroscopic analysis on the causes of traffic congestion. Based on Traffic Performance Index (TPI), a dynamic macroscopic index showing the whole area congestion intensity.

Keywords: *Down Town Streets, Traffic Congestion, Capacity & Level Of Service, Multiple Regration Equation.*

INTRODUCTION:

TRAFFIC IN HYDERABAD CITY

Due to rapid urbanization, the tremendous rise in number of vehicles variably accompanied by ever increasing volume of traffic and causes of traffic congestion on road almost every city in India facing acute traffic congestion, delay ,pollution , accidents etc.

One of the main sources of the increased emission load from vehicles is the growing traffic congestion in the city. Evidence from numerous cases around the globe suggests that effective traffic management measures dealing with this problem should necessarily include both supply and demand based approaches. However, the existing traffic measures in Hyderabad only focus on the supply side: strategies that improve the physical aspects of roads by increasing the road capacity either by widening existing roads or by constructing new roads or flyovers.

Ideally, urban transport policies should be developed on the basis that congestion is related to both:

- The behavior of traffic as it nears the physical capacity of the road system.

S.No: (2)

TRAFFIC MANAGEMENT IN URBAN AREA

MORA MADHU
M Tech Student,
Mail:madhumora1991@gmail.com

Y VARUN KUMAR
M Tech Student,
Mail:varunchary12@gmail.com

DR. R. RAMESH REDDY
(Professor),
Mail:rameshlalitha@yahoo.com

ABSTRACT

Area traffic management is an essential prerequisite for the effective planning in urban area. Most of the traffic problems are caused by certain deficiencies in planning, engineering and management issues. The traffic congestion, vehicular delays, risk of accident, uneconomic travel and other psychological strains are the observed features in the urban area.

The development of built up areas is a reflection of land use change in urban activated areas. The road infrastructure density, its functionality, hierarchical connectivity should reciprocate to the demand profiles like trip length, trip interactions and trip densities. An approach for identification of urban longitudinal and transitional corridors through the development of primary road system is suggested in this study. This approach suggests the decongestion measures in mixed traffic conditions by identifying the operational routes for the external based trips. The relational patterns with traveler characteristics (which are a reflection of land use density) are used for developing a policy on corridor identification and cost optimization in travel. This will also promote a way for environment sustenance in mixed traffic, improper land use development and non-planned road network conditions.

A lead on user average trip lengths, trip orientation and trip intensity is obtained from the travel surveys and its successive analysis. The data used in the study is obtained by the comprehensive surveys like Traffic studies, Video graphic surveys, Road densities, road hierarchical connectivity and road functionality have been ascertained from the source of satellite data and the relevant field surveys. The analysis has given a lead for identifying the corridors needed for development in urban areas. The planning is made based on the concepts of user preferences, static analysis of land use , road length and distribution of access points, trip behavioral patterns, future travel demand , static – dynamic correlations and prioritization of routes.

This project envisages studying the various issues related to traffic congestion like level of service, speed, congestion, pedestrian, facility in a given local area. The study of including congestion survey to capture the traffic data in a corridor covering from TV tower to Kothapet on the NH-65 in Hyderabad.

From the survey data the level of service on the street was found out an indicates level of "D" indicating high congestion the facility for parking of vehicles were in adequate unauthorized parking at many location was identification proper walkways for pedestrians were absent and pedestrians crossings were also not provided. The observation to traffic due to unauthorized encroachment was observed.

Having identified the deficiencies from the analysis of the data from survey, recommendations area made to improve the condition by necessary intersection.

INTRODUCTION

PROBLEM CONTEXT:

The rapid growth of urban population generates problems like congestion with increase in traffic, unbalanced land use pattern and its distribution under different land users, growth of slums in the core as well as at the peripheries of the urban areas and degradation of environment etc. in order to bring a balanced urban growth, a proper understanding of geographic characteristics, land use distribution and population levels predicted to a future date may help on urban planner to develop sound and rational planning methodologies for solving the urban problems.

S.No: ③ ,

EFFECT OF PATTERN LOADING ON SEISMIC BEHAVIOUR OF STRUCTURES BY CONSIDERING INFILLED EFFECT

R.JAIPAL
M.Tech Student, Department of Civil,
Bharath Institute of Engineering &
Technology, Hyderabad, India, 500070
E-mail: jaipalcivil.jntu@gmail.com

Y.VARUN KUMAR
Assistant Professor ,Department of Civil,
Annamacharya Institute of Technology &
Sciences, Hyderabad, India, 501512
Email: varunchary12@gmail.com

ABSTRACT
The accurate estimation of design actions on the structure is very important in structural design as it significantly affects the final design and objectives. Any error in the estimation of design actions may lead to wrong results of structural analysis on the structure and lead to the unrealistic sizing of its structural members or even collapse of the structure. Therefore it is important to account for the most adverse effects of live loads on the structure. The consideration of pattern loading depends on the ratio of dead to live load and the type of structural member.

These days most of the engineers are not considering the different live load patterns to get the adverse effect of the structure. Considering the live load to all the slab panels may not appropriate to estimate the design parameters. In this context, an attempt is required to see the effect of pattern live load on the structure under seismic loads. The effect of pattern load may be different from bare frame structure and also infill structures. For the present work a regular symmetrical building will be chosen and the structure is loaded with different pattern live loading is analyzed for seismic load case with and without infill walls. Different dead loads to live load ratios are also considered as a parameter.

Index Terms—Live load patterns, pattern loading, infill walls

I. INTRODUCTION

Generally the structural deisgn of individual members are designed for the critical values of analysis results. Identifying worst analysis results estimation with different load combinations are crucial for any designer to avoid the failure of members or structure as whole. The loads on any structure are fixed or movable. Self weight of members are in fixed loads category where as live loads such as human occupancy floor loads can be placed in various ways are in the category of movable loads. The live loads position have influence on behaviour of structure. Hence live load position need to be consider in analysing the sturctures. such live load positions are known as live load pattern. Assuming the patterns of live load for worst response of the stucture is more crucial and difficult in multi dimensional systems. Such situations required number of trails to be attempted. Conventionally dead loads, live loads, earthquake loads and wind loads are the primary load types used to analyze a structure for various parameters like span moments, end moments, shear, thrust or deflections. The Muller Breslau Principle for influence lines is an effective way to obtain critical load patterns. Realizing the fact that the efforts required in solving large structures is too much and such efforts further increase as design demands multiple analysis of the structure. In a way, such conventional analysis tools prove to be realistic only in a qualitative sense. Further, combining load combinations and load patterns requires the engineer to do multiple iterations of structural analyses in order to capture the critical scenario. Apart

from being an impractical task in most situations, it is impossible at times. In fact for Simplicity standard structural engineering codes of practice have suggested several critical load patterns.

The objective of present study is to understand the behaviour of a symmetrical building of G+9 stories under static and seismic loads for different live load patterns. It is to check importance of considering live load patterns for the analysis of structures.

II. LITERATURE REVIEW

ASCE 7-05 Section 4.6 states "The full intensity of the appropriately reduced live load applied only to a portion of a structure or member shall be accounted for if it produces a more unfavorable effect than the same intensity applied over the full structure or member." What this means is that it is need to arrange the live load so as to cause maximum effect in members. The design of structural elements must have sufficient strength to support all possible arrangements of live load. Consequently the analysis needs to provide with envelope diagrams for each member. Envelope diagrams are internal force diagrams that envelop all the possible values of force at each location along the member. So examples are used below to explain method for determining envelopes. This can seem daunting task as designer need to do multiple load cases to account for the various loadings on your structural system. For statically determinate structures, it is often easy to establish critical loading scenarios for shear, moment, reactions, and deflection. Unfortunately for continuous, statically indeterminate structures this is not so obvious and the use of influence lines becomes extremely useful.

Furlong R.W (1981) worked on rational analsysis of multistory concrete structures. Problems to be faced in

S.No: (4)

A CASE STUDY ON PARKING AND ITS MANAGEMENT IN HIGH DENSED URBAN CORRIDORS

Mr. AVINASH REDDY
M Tech - Transportation Engineering
Vishwabharathi College Of Engineering-Hyderabad.

Mr. CHINNAM TILAK
M Tech – Transportation Engineering
Vishwabharathi College Of Engineering-Hyderabad.

ABSTRACT

The number of vehicles is increasing at an alarming rate in the urban areas. The commercialization of the area is also happening at a fast rate. The investment on roads and for parking facilities have not kept in pace with these growing traffic leading to congestion and accidents. The propensity to own private vehicles and the necessity for their use has generated huge parking demand in metropolitan cities. Almost all the metropolitan cities are experiencing increased problems related to parking. With the rapid increase of cars the need to find available parking space in the most efficient manner, to avoid traffic congestion in a parking area, is becoming a necessity in 'car park management'.

Many cities in the developing world are rapidly growing and the economic patterns of the people living in these cities in changing. With these changes there is a need for these cities to stay up to the mark in providing the mobility facilities or in other words meet the needs of mobility for the citizens. Often city officials presume that the providing of more parking spaces for the citizens means meeting the mobility needs. On the contrary, every car that is on the road needs a place to be parked. It is a key issue in almost all urban areas.

One of the problems created by road traffic is 'parking'. Not only do vehicles require street space to move about, but also they do require space to park where the occupants can be loaded and unloaded. The period over which a car is parked is very great compared with the time it is in motion. The size of average parking is 14m^2. It is roughly estimated that out of 8760 hours in a year, the car runs on an average for only 400 hours, leaving 8360 hours when it is parked. Every car owner wish to park their car as closely as possible to his destination so as to minimize walking.

Cars take up space when they are moving but for an average of 23 hours of the day they are parked, and if they were to be used for all journeys then they would need a parking space at both ends of every trip – so many spaces are required for every car. A parked car takes up around 8 square meters when parked and often the same again in maneuvering space – a huge amount in dense urban areas where land is expensive. Often, cars get more space to park than humans have to live in! The above mentioned reason justifies the need for having a parking management system.

More focus needs to be devoted towards better public transport and non-motorized transportation. Parking needs to be used as a demand management tool. Transportation experts recognized that 'Parking' is a critical component of Transportation Policy and Management as it effects travel behavior, safety, economic development, revenue, land use, traffic congestion and air quality etc. 'Parking management' refers to various policies and

S.No: (5), (6)

International Journal of
Advanced Technology and Innovative Research

www.ijatir.org

ISSN 2348–2370
Vol.09,Issue.12,
November-2017,
Pages:2218-2224

Effect of Stone Dust and Fines on the Properties of High Strength Self-Compacting Material

Y. VARUN KUMAR[1], R. JAIPAL[2], SIDDHI RAMULU[3]

[1]PG Scholar, Dept of Civil, Siddhartha Institute of Technology & Sciences, Ibrahimpatnam, Hyderabad, TS, India.
[2]Assistant Professor, Dept of Civil, Siddhartha Institute of Technology & Sciences, Ibrahimpatnam, Hyderabad, TS, India.
[3]Assistant Professor, Dept of Civil, Siddhartha Institute of Technology & Sciences, Ibrahimpatnam, Hyderabad, TS, India.

Abstract: Self-compacting concrete (SCC) represents a milestone in concrete research. SCC is a highly flow able, non-segregating concrete that can spread in to place, fill the formwork and enclose the reinforcement without any mechanical vibration for consolidation. Concrete technology has made for significant advances in recent years which results in economical improvement of the strength of concrete. This economical development depends upon the intelligent usage of the locally available materials. Important constituent of self-compacting concrete (SCC) is natural sand and filler material which is expensive and rare. This necessitates that a suitable substitute be found. The cheapest substitute for natural sand is quarry dust and for filler material is fly ash. Quarry dust, a by-product from the crushing process during quarrying activities is one of the materials being studied and fly ash is an artificial pozzolanic material, a finely divided pozzolana form compounds which have cementitious properties, when mixed with hydrated lime and alkalies. In this work, the fresh split tensile and compressive strength properties of self-compacting concrete when the sand is partially replaced with stone dust , when the filler materials is increased by adding fly ash in % of the total powder content and when both substituent's are implemented simultaneously. Optimization of stone dust and fly ash is also obtained. The results indicated that the inclusion of quarry dust into the self-compacting Concrete mix as partial replacement material to natural sand resulted in higher compressive strength, low tensile strength and optimization of sand replacement is 40%. Optimization of addition of fly ash in total powder content is 30%.

Keywords: Stone Dust, Workability Tests, Compression Test, FlyAsh, Tensile Strength.

I. INTRODUCTION

Self-compacting concrete was first developed in 1988 for improve durability of concrete structures. Since then, various investigations have been carried out and this concrete is first used in practical structures in Japan, mainly by large construction companies. To make this a standard concrete several rational mix-design methods and self-compact ability testing methods have been carried out. Development of self-compacting concrete is a wished achievement in the construction industry in order to overcome problems associated with cast in place concrete. Compared to normally vibrated concrete (NC), self-compacting concrete (SCC) possesses enhanced qualities and improves productivity and working conditions due to the elimination of compaction. In order to achieve optimum strength and durable concrete structures compaction is the key. But full compaction was difficult to obtain because of increasing in reinforcement volumes with smaller bar diameters and a reduction in skilled construction workers, leading to poor quality concrete. Self-compacting concrete is not affected by the skills of workers, the shape and amount of reinforcing bars or the arrangement of a structure and also due to its high fluidity and resistance to segregation it can be pumped longer distances. Performance concrete having compatibility in fresh stage and no initial defects in early stage.

The concept of self-compacting concrete was first proposed in 1986 by professor hajimeokamura (1997), but the prototype was first developed in 1988 in Japan, by professor ozawa (1989)at the university of Tokyo. Recommendations on the design and applications of SCC in construction have been developed by many professional societies, including the American Concrete Institute (ACI), the American Society for Testing and Materials(ASTM), Centre for Advanced Cement-Based Materials (ACBM), Precast Consulting Services (PCI) and Reunion International Laboratories et Experts des Matériaux, systèmes de construction et outrages (RILEM) etc. Self-compacting concrete is direct so that no of additional inner or outer vibration is necessary for the compaction. It flows like honey and has a very smooth surface level after placing. The composition of SCC is similar to that of normal concrete but to attain self flow ability some chemical and mineral admixtures are used. Usually, the chemical admixtures used are high range water reducers (super plasticizers) and viscous modifying agents (VMA), which change the rheological properties of concrete.

II. LITERATURE REVIEW

Self-compacting concrete (SCC) represents one of the most significant advances in concrete technology for

S.No:- (8)

ANALYSIS AND DESIGN OF A HEAVY CARGO BERTHING STRUCTURE

[1]B.RAGHAVA MAHEEDHAR, [2]C.V.SIVA RAMA PRASAD
[1,2]Assistant Professor
[1,2]Department of Civil Engineering,
[1]Annamacharya institute of Technology and sciences,Piglipur,Batasingaram(V),Hayatnagar (M),R.R.Dist-501512,India
[2]Vignana bharathi institute of technology, Aushapur(V), Ghatkesar(M),R.R.Dist-501301,Hyderabad,India

Abstract— The structures which are constructed for the intention of berthing and mooring of vessels to facilitate loading and unloading of cargo and also for embarking and disembarking of passengers or vehicles etc. is called berthing structure. Various factors influence the analysis and design of the berthing structures. The berthing structures are designed for dead load, live load, berthing force, mooring force, earthquake load and other environmental loading due to winds, waves, currents etc. In the present study, a proposed berthing structure EQ-10 is taken for analysis and design .All suitable data is collected from Visakhapatnam port trust and their website like geotechnical data, environmental data, and traffic forecasting data. By using all these data, we planned and modeled a structure. After that we calculated various loads induced on structure and we analyzed the modeled structure in STAAD-PRO due to the typical load distribution on structure. Actually we have trailed with different dimensions for most acceptable structure, in that trailing we concluded that larger diameter pile gets less deflection when compare with smaller diameter piles. Finally the structure was analyzed and designed with resisting of marine conditions and satisfying in the aspect of economical and safety

Key words—Berthing structure, STAAD-PRO, Marine Conditions.

I. Introduction

In the present study, we described a suitable way to design a new berthing structure with example of one of the proposed berthing structure in Visakhapatnam port. So before analyzing and designing, the influence factors which effected on the structure were taken into consideration such as soil characteristics of the proposed location, environmental conditions and range of traffic. All the basic Data was adopted from Visakhapatnam port which were supposed to be used in the project such as geotechnical data, environmental data, and traffic forecasting data. The entire Berth length of 100m was divided into 3 units of each 33.33 in length with an expansion joint of 40mm between successive units and proposed in the inner harbor, meant for handling liquid cargo like Sulphuric acid, Phosphoric acid, phosphoric acid, edible oils etc. The details of the structural element are discussed under the conceptual design. The design dredge level is taken as -16.10m. Factors to be considered before going to design a berthing structure like fixing of a location, selection of type of berth, deciding of Number of berths, selecting Length of berth and Area of berth, required Draft alongside berth ,Apron width, Deck elevation, turning circle, and Stacking area requirements Area requirements for other facilities. The entire EQ (Eastern Quay)-10 berth length of 100.07 m is divided into 3 units of each 33.33 in length with an expansion joint of 40mm between successive units. The proposed EQ-10 berth at Visakhapatnam Port in the inner harbor is meant for handling liquid cargo like Sulphuric acid, Phosphoric acid, phosphoric acid, edible oils etc. the details of the structural element are discussed under the conceptual design .although the concession agreement provides for dredging has to be carried up to -16.10m .hence the design dredge level is taken as -16.10m

II. GEOMETRY OF STRUCTURE

Thickness of apron layer	: 200mm
Thickness of slab	: 300mm
Size of transverse beam	: 1800mmX1100mm
Size of longitudinal beam	: 1100mmX600mm
Size of pile	: 1.70 diameters, height 21.65 meters
Total height of the structure	: 23.30meters
Design dredged level	: 16.60 meters
Pile submerged level	: 19.60 meters
Deck elevation	: 3.70mt
Kerb wall height	: 1mt
Area of berth	: 100m X12m
Number of divided units	: 3
Area of each unit	: 33m X 12m
Slab panel size	: 2.62m X 2.62m

III. LOADS ON STRUCTURE

Wearing coat (Apron)	= 5 kN/m^2(density of the concrete is taken 25 kN/m^3)
Slab weight	= 7.55 kN/m^2
Beams	
Transverse beams	= 50kN/m
Longitudinal beam	= 16.5 kN/m
Pile	= 920.12 kN/m

Live load is based functioning of berth and truck loading on berth as per IS: 4851 (Part III) – 1974. The function of berth related to Truck loading A or AA or 70R (Heavy cargo berth) so we are adopted 50 kN/m^2.

S.No. (9), (10)

COMPRESSIVE STRENGTH STUDY ON GLASS POWDER REPLACEMENT WITH FINE AGGREGATE

[1]M.ARUN KUMAR, [2] B.RAGHAVA MAHEEDHAR
[1,2]Assistant Professor
[1,2]Department of Civil Engineering,
[12]Annamacharya institute of Technology and sciences, Piglipur, Batasingaram (V), Hayatnagar (M), R.R.Dist-501512, India

Abstract— *Glass powder (GP) used in concrete making leads to greener surroundings. In glass industries, broken glass sheets & sheet glass cuttings are go to waste, which are not recycled at present and usually added to landfills for disposal. The use of GP in concrete is an interesting possibility for economy system on waste disposal sites and conservation of surroundings. This project examines the opportunity of the usage of GP as fine aggregates replacement in concrete. Natural sand changed into partly changed (25%, 50%, 75%, 100%) with GP in concrete. Compressive strength up to 28 days of age had been as compared with the ones of high performance concrete made with natural sand.*

key words—Glass powder, compressive strength, economy system.

I. INTRODUCTION

As modern engineering practices become more demanding there is a corresponding need for special types of materials with novel properties. Scientists, Engineers and technologists are continuously on the lookout for materials, which can act as substitute for conventional materials or which possess such properties as would enable new designs and innovations resulting into economy, so that a structure can be built economically.

However on many occasions individual materials as such may not serve the specific purpose. There have been so far many attempts to develop new materials, which is the combination of two or more materials. Such materials are called Composite material.

Concrete can be regarded as a composite material. For reducing the cost of concrete, greater use of pozzolanic materials like fly ash, blast furnace slag and waste glass was suggested. The use of these materials as the substitute material in concrete would reduce the disposal problem faced by thermal power plants and industrial plants and at the same time achieving the required strength of concrete.

Glass is amorphous solid material which is produced at high temperatures followed by crystallization. The effective use of waste glass for partial and full replacement of sand as an admixture in cement mortar and concrete has established in the country in recent years.

Recent investigation of waste glass has indicated greater scope for their utilization as a construction material. Greater utilization of waste glass will lead to not only saving such construction material but also assists in solving the problem of disposal of this waste product.

The recent investigations have also indicated the necessity to provide proper collection methods for waste glass so as to yield waste glass of quality and uniformity, which are primer requirements as waste glass for use as construction.

II. MATERIALS TEST RESULTS

Table1: Physical properties of glass powder

S.No	Physical properties of glass powder	
1.	Specific gravity	2.58
2.	Fineness Passing 150μ	99.5
3.	Fineness Passing 90μ	98

Table2: Chemical Properties of glass powder

S.No	Chemical Properties of glass powder	
1.	pH	10.25
2.	Colour	Grayish White

Table3:Chemical Properties of glass powder

S .NO	Chemical properties of glass powder	% By Mass
1.	SiO_2	67.330
2.	$Al2O_3$	2.620
3.	Fe_2O3	1.420
4.	TiO_2	0.157
5.	CaO	12.450
6.	MgO	2.738
7.	Na_2 O	12.050
8.	K_2O	0.638

S.No: (11), (13), (14) Report.

RESEARCH ARTICLE OPEN ACCESS

COMPARATIVE STUDY ON ANALYSIS AND DESIGN OF G+ 12 STOREYS BUILDING WITH AND WITHOUT BASEMENT WALLS OF TWO BASEMENTS

, M.Arun kumar [1], B.Raghava Maheedhar[2] G.Janakiram goud[3]

1(Civil Engineering, Annamacharya institute of Technology and sciences, Piglipur, Batasingaram (V), Hayatnagar (M), R.R.Dist-501512, India .)

2 (Civil Engineering, Annamacharya institute of Technology and sciences, Piglipur, Batasingaram (V), Hayatnagar (M), R.R.Dist-501512, India .)

3 (Civil Engineering, SS ventures private limited, Kukatpally, R.R.Dist-500072, India.Email)

Abstract:

The earthquake phenomenon represents one of the maximum devastating forces that reasons no longer only loss to human lifestyles but cripples the economic system of a country as properly. Hence this project work is aimed to study the effect of basements with and without basement walls on seismic behaviour of multi storey building. In the present study seismic analysis has been done for a G+12 storey buildings with and without basement walls of two basements by using Strap (structural analysis programme) software.

Keywords — **Basements with & without basement walls, Seismic behaviour of multi storey building, G+12 storey building, Strap software.**

I. INTRODUCTION

A basement or cellar is one or more floors of a building that are either consummately or partially below the ground floor. The word cellar or cellars is utilized to apply to the whole underground level or to any sizably voluminous underground room. A sub cellar is a cellar that lies further underneath. A basement can be utilized in virtually precisely the same manner as an adscititious above-ground floor of a house or other building. However, the utilization of basements depends largely on factors categorical to a particular geographical area such as climate, soil, seismic activity, building technology, and authentic estate economics. There has been a growing trend to construct basements in the expedient of elongating accommodation and parking. Basements are additionally built as a component of both incipient residential and commercial developments however this is not an incipient trend. Recently, most of the high-elevate buildings may have basements utilized as parking lots or shopping malls etc. In general, it is commonly surmised that the building is fine-tuned at the ground level in the analysis and the basement is not included in the analytical model when the basement walls are connected to the floor deck and in between columns. Utilizing this posit, the natural periods may be abbreviated due to the flexibility introduced by the basements.

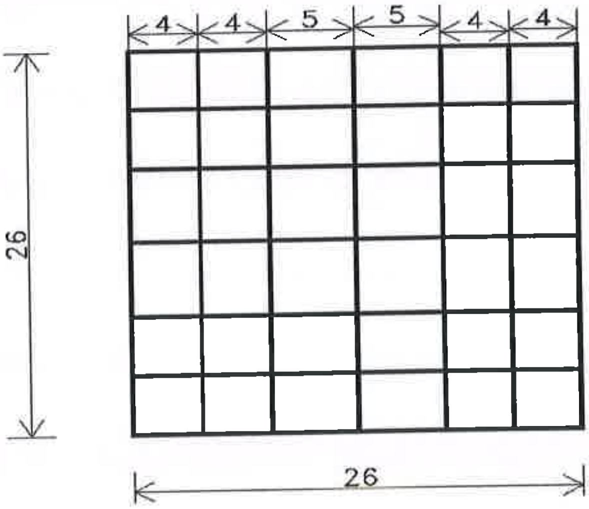


Fig 1: Layout of building (All dimensions in meters)

PRINCIPAL
Annamacharya Institute of
Technology & Sciences
Piglipur (V), Batasingaram (Post),
Abdullapurmet (M), R.R.Dist.HYD-501 512.

ANNAMACHARYA INSTITUTE OF TECHNOLOGY & SCIENCES * PIGLIPUR (V), R.R.Dist. HYD. ESTD: 2005

S.No: 12, 15

RESEARCH ARTICLE OPEN ACCESS

A Comparative Study on Various Mix Designs of Concrete by Using Steel and Glass Fiber

B.Raghava Maheedhar[1], M.Arun kumar[2], C.V.Siva Rama Prasad[3]

1(Civil Engineering, Annamacharya institute of Technology and sciences, Piglipur, Batasingaram (V), Hayatnagar (M), R.R.Dist-501512, India.)

2 (Civil Engineering, Annamacharya institute of Technology and sciences, Piglipur, Batasingaram (V), Hayatnagar (M), R.R.Dist-501512, India)

3 (Civil Engineering, Vignana bharathi institute of technology, Aushapur(V), Ghatkesar(M),R.R.Dist-501301,Hyderabad,India.)

Abstract:

There is always a search for concrete with higher strength and durability. Plain concrete has good compressive strength but has low tensile strength, low ductility and low fire resistance. This research paper aim to study characteristics and comparison of the mechanical properties of steel and glass fiber reinforce concrete with conventional concrete. In order to achieve and verify that 1%,2%,3% fiber percentage by the volume of cement are used in this study with three different concrete mixes M20, M25, and M30. 7 and 28days compressive strength and 28 days spilt tensile strength and flexural strength; tests have been performed in the hardened state. In this project the behaviour of cube, cylinder & beam structures strengthen by using FRC is experimentally tested. The fiber used are steel and glass fibers in sundry volume fraction the main reason for integrating steel fiber to concrete matrix is to ameliorate the post cracking replication of the concrete i.e. to ameliorate its energy absorption capacity and ostensible ductility and to provide a crack resistance and crack control and addition of glass fiber form bridging the micro-cracks are suggested as the reason for the enhancement in flexural strength.

Keywords — **Steel fiber, Glass fiber, Compressive strength, Split tensile strength, flexural strength.**

I. INTRODUCTION

Concrete is considered a brittle material, primarily because of its low tensile strain capacity and poor fracture toughness. Reinforcement of concrete with short randomly distributed fibers can address some of the concerns related to concrete brittleness and poor resistance to crack growth. Fibers, used as reinforcement, can be effective in arresting cracks at both micro and macro-levels. At the micro-level, fibers inhibit the initiation and growth of cracks, and after the micro-cracks converts into macro-cracks, fibers provide mechanisms that abate their unstable propagation, provide effective bridging, and impart sources of strength gain, toughness and ductility. Concrete can be modified to perform in a more ductile form by the addition of randomly distributed discrete fibers in the concrete matrix.

Certain disadvantages like brittleness and poor resistance to crack opening and spread. Concrete is brittle by nature and possess very low tensile strength and therefore fibers are used in one form or another to increase its tensile strength and decrease the brittle behaviour. With time a lot of experiments have been done to enhance the properties of concrete both in fresh state as well as hardened state. The basic materials remain the same but super plasticizers, admixtures, micro fillers are also being used to get the desired properties like workability, Increase or decrease in setting time and higher compressive strength.

FRC can be regarded as a composite material with two phases in which concrete represents the matrix phase and the fiber constitutes the inclusion phase. Volume fraction of fiber inclusion is the most commonly used parameter attributed to the properties of FRC. Fiber count, fiber specific surface area, and fiber spacing are other parameters, S.No: 16 17

Analysis and Design of g+12 storey building with shear wall effect with two basements

B.Raghava Maheedhar[1], M.Arun Kumar[2], S.Nagarjuna [3], C.V. Siva Rama Prasad[4]

[1,2,3] Assistant Professor, Department of civil engineering, Annamacharya institute of technology and sciences, Piglipur, Batasingaram (V), Hayatnagar (M), R.R.Dist-501512, Hyderabad, India.
[4] Assistant Professor, Department of civil engineering, Vignana bharathi institute of technology, Aushapur(V),Ghatkesar(M),R.R.Dist-501301,Hyderabad,India.

ABSTRACT - *Now a day's most of the earth quake resistant buildings are provided with shear walls hence I included shear wall in my model. The main objects of this study were to investigate the behavior of multi storey building with and without basement walls with shear wall effect. For the study two models of G+12 storey buildings with two basements along with shear wall are considered. The building has six bays in X1 direction and six bays in X3 direction with the plan dimension of 26m × 26m. The constructing is kept symmetric in both orthogonal instructions in plan to avoid torsional reaction. Underneath pure lateral forces the orientation and length of columns is kept same throughout the height of the structure.*

Key Words: earth quake resistant buildings, multi storey building, multi storey building with basement walls with shear wall effect, multi storey building without basement walls with shear wall effect

1. INTRODUCTION:

When a structure is subjected to ground motions in Associate in nursing earthquake, it responds by moving. The random motions of the bottom caused by Associate in nursing earthquake are often resolved in any 3 reciprocally Perpendicular directions: the 2 horizontal directions (x and z) and also the vertical direction (y). This motion causes the structure to vibrate or shake all told 3 directions; the predominant direction of shaking is horizontal. All the structures square measure primarily designed for gravity hundreds force adequate to mass time's gravity within the vertical direction. Thanks to the inherent issue of safety employed in the look specifications, most structures tend to be adequately protected against vertical shaking. Generally, however, the inertia forces generated by the horizontal parts of ground motion need bigger thought in seismic style. Earthquake generated vertical inertia force should be thought-about within the style unless checked and proved to be insignificant, In general, buildings aren't notably liable to vertical ground motion, however its result ought to be borne in mind within the style of RCC columns, steel column connections, and pre-stressed beams. Vertical acceleration ought to even be thought-about in structures with massive spans, those within which stability may be a criterion for style, or for overall stability analysis of structures with massive spans. Structures designed just for vertical shaking, in general, might not be able to safely sustain the result of horizontal shaking. Hence, it's necessary to confirm that the structure is sufficiently immune to horizontal earthquake shaking too.

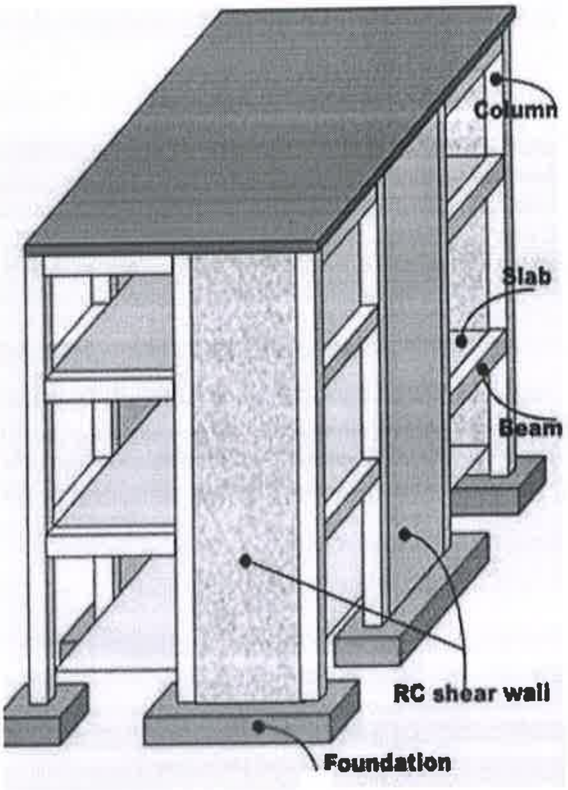


Fig -1: Reinforced shear walls in buildings

Table -1: Important features of building

1	Type of structure	Multi storey special moment resisting frame
2	Zone	3
3	Layout	As shown in fig 6.1
4	Number of stories	G+12 storey building with shear wall
5	Number of basements	2
6	Floor to floor height	3 m
7	External walls	230 mm
8	Internal walls	150 mm

S.No: 18

Intensification of Non-Breakable Material Ray with Glass Fiber Reinforced Polymer

Jeedi Vennela[1], Dr.MD.Subhan[2]
[1]M.Tech Student, Dept of Civil Engineering, [2]Professor & HOD, Dept of Civil Engineering
[12]AVN College of Engineering and Technology, Hyderabad, T.S, India.

Abstract- Experimental investigations on the flexural and shear behaviour of RC beams strengthened using continuous glass fiber reinforced polymer (GFRP) sheets are carried out. Externally reinforced concrete beams with epoxy-bonded GFRP sheets were tested to failure using a symmetrical two point concentrated static loading system. Two sets of beams were casted for this experimental test program. In SET I three beams weak in flexure were casted, out of which one is controlled beam and other two beams were strengthened using continuous glass fiber reinforced polymer (GFRP) sheets in flexure. In SET II three beams weak in shear were casted, out of which one is the controlled beam and other two beams were strengthened using continuous glass fiber reinforced polymer (GFRP) sheets in shear. The strengthening of the beams is done with different amount and configuration of GFRP sheets.

Experimental data on load, deflection and failure modes of each of the beams were obtained. The detail procedure and application of GFRP sheets for strengthening of RC beams is also included. The effect of number of GFRP layers and its orientation on ultimate load carrying capacity and failure mode of the beams are investigated.

Keywords- GFRP; flexure; shear; strengthened; symmetrical two point concentrated static loading system.

I. INTRODUCTION

The rehabilitation of infrastructures is not new, and various projects have been carried our around the world over the past two decades. One of the techniques used to strengthen existing reinforced members involves external bonding of steel plates by means of two-component epoxy adhesives. It is possible to improve the mechanical performance of a member. The wide use of this method for various structures, including building and brides, has demonstrated its efficiency and its convenience. In spite of this fact, the plate bonding technique presents some disadvantages due to the use of steel as strengthening material. The principal drawbacks of steel are its high weight which causes difficulties in handling the plates on site and its vulnerability against corrosive environments. Moreover, steel plates have limited delivery lengths and, therefore, they require joints.

II. RELATED STUDY

In this report three beams were tested for flexure, controlled beam and other two beams were casted and strengthened by applying GFRP on two beams in flexure mode. Further study continues by testing more number of beams includes various kinds of fiber materials. A further study includes the strengthening of beam is done by different amount and different configurations of GFRP sheets provided. The various concrete mix proportions and also cross sectional dimensions of the beam and analysis also consider for further study. Strengthening of the beam is also depends on matrix materials like epoxy resin (adhesives). The matrix materials have mechanical properties such as strength, shear and compression. So we have a scope for further study that by using different kind of matrix materials to strengthen the reinforced concrete beams.

III. METHODOLOGY AND TASTING

Two sets of beams were casted for this experimental test program. In SET I three beam (F1, F2 and F3) weak in flexure were casted using same grade of concrete and reinforcement detailing. In SET II there beams (S1, S2 and S3) weak in shear casted using same grade of concrete and reinforcement detailing. The dimensions of all the specimens are identical. The cross sectional dimensions of the both the set of beams is 150 mm by 150 mm and length is 700 mm. in SET I beams 2, 10 mm ϕ bars are provided as the main longitudinal reinforcement and 6 mm ϕ bars as stirrups at a spacing of 100 mm center to center where as in SET II beams 3, 10mm ϕ bars are provided as the main longitudinal reinforcement and without any stirrups.

A. ERFORMANCE BASED OBJECTIVE

An objective of performance based objective targets like the flexural behavior of reinforces concrete beams. To study the effect of GFRP strengthening on ultimate load carrying capacity and failure pattern of reinforced concrete beams. Another objective is based on the shear behavior of reinforced concrete beams. To study the effect of GFRP strengthening on the shear behavior of reinforce concrete beams.

TESTING:

The flexural and shear strength of a section depends on the controlling failure mode. The following flexural and shear failure modes should be investigated for an FRP strengthened section.

S.No:- (19)

SPEED, FLOW AND HEADWAY MODELING OF URBAN TRAFFIC IN HYDERABAD CITY AT DIFFERENT LOCATIONS

[1]B.CHITTI BABU , , M.Tech Scholar

[2]K.APARNA, Assistant Professor

[3]Dr. ANAND SWAROOP GOYAL , Professor

DEPARTMENT OF CIVIL ENGINEERING

Ashoka Institute of Science & Technology, Hyderabad, Telangana, India

ABSTRACT

Issues related to mixed traffic conditions are unique on the urban roads of developed countries. Persists, bicycles, buses, cars, motorcycles / scooters, auto barriers, bicycle racks and various types of travel methods have created non-communication partnerships on the same street location, which is the economic potential of the cities in developed countries. Coming back there is no problem solving these issues for different reasons. Failure to solve these problems is making them more complicated. In developing countries, there is to increase the increase in motor vehicle property. Therefore, the information about the basic traffic flow features and the associated analytical technologies in the planning, design and operation of transport systems is necessary. The principle of traffic flow makes traffic engineer so that the flow between traffic flow, density and speed on the road. Many researchers have suggested that in connection with the characteristics of traffic flow on the roads. The main objective of this study is to develop mathematical models for the conditions of different types of flow and finding suitable headway in this study, for speed, flow, and front traffic conditions. The study section estimates the ability to speed and flow relationships.

I. INTRODUCTION

1.1 GENERAL

The problems associated with mixed traffic conditions on city streets of growing international locations are specific. Pedestrians, bicycles, buses, motors, motorcycles/scooters, auto rickshaws, cycle rickshaws and various different forms of journey modes share the equal street area developing inefficient mobility situations which are robbing the economic capacity of the cities in developing international locations. Because of numerous motives these troubles are yet to be resolved. The incapacity to clear up those issues is making them increasingly more complex. Increasing motor car possession inside the growing international locations is further annoying the scenario. Consequently know-how of fundamental traffic glide traits and associated analytical strategies is an essential requirement inside the planning, layout and operation of transportation structures.

1.2 URBON TRANSPORT SCINORIO OF IN INDIA

City delivery is one of the maximum vital components of urban infrastructure. A very good community of roads and a green Mass city shipping device make a significant contribution to the “running efficiency” of a large town. A poor city delivery system can also gradual down financial boom of the town and also lead to its decay. It has been predicted that the bad traffic and Transportation scenario in the urban areas of the use currently result in an annual loss of the order of Rs. 20,000 crores in motors operating and journey time charges by me. In view of the hastily growing city populace stress on urban shipping device is certain to growth much greater inside the coming years. Pressing measures are therefore had to tackle this problem.

On average for the duration of peak hours in Mumbai the actual occupancy in a suburban educate is in extra of 4000 passengers that have maximum desirable capacity of 2600 passengers. Most of the Indian cities have greater or much less comparable visitor’s congestion. Estimates for the metropolitan towns show that approximately eighty million journeys will need to be catered to consistent with day while most effective 37 million journeys are being furnished by way of the available rail and bus mass delivery facilities. Moreover in line with an international financial institution look at for every extra a million people in a growing town a further 3.5 to four million

S.No: 20

International Journal of Research
Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue-17
December 2017

A STUDY AND ANALYSIS OF CAPACITY AT UN SIGNALIZED INTERESTIONS

Mr.PATTA RAJESWARA RAO ,PG STUDENT
DR.ANAND SWARUP GOYAL (HOD), ASSOCIATE PROFESSOR
UDAY KIRAN (M.Tech) , Asst.professor

Department of Civil Engineering
ASHOKA Institute of Engineering & Technology(Approved By AICTE & Affiliated to JNTU. Hyd)
Malkapur (V). Choutuppal (M). Nalgonda Dist-508252

Abstract:

Unsignalized intersection is implemented to regulate low volume of traffic flow. The gap-acceptance method is the common approach to assess the performance of the intersection. However, among the drawbacks of the gapacceptance method are the non-compliance to the right of way, and the heterogeneous traffic condition. Conflict method is developed to overcome these shortcomings. Surveillance equipment is used to obtain the required data, such as traffic volume and occupation time. The occupation time of vehicle is used to calculate the capacity of vehicular movements for each conflict group. The control delay and level of service of the vehicular streams are evaluated according to the procedures in HCM 2000. Result comparison is made between the conflict method and the HCM 2000. The relationship between the occupation time and critical gap is discovered. The results of the conflict method are found to be comparable with the HCM 2000 using field data.

I.INTRODUCTION

An intersection is a node, and usually it is a bottleneck for traffic flow in highway network. Capacity of a intersection affects the total capacity of highway network due to all types of turning movements. For actions of conflicting, merging and diverging caused by traffic flow, the traffic characteristics of intersection are more complex than those of road mid block section. Traffic stream in developing countries comprises of different types of motorized and non-motorized vehicles leads to mixed traffic conditions and lane changing patterns.

Urban roads in India carry different types of vehicles like high speed automobiles, low speed cycles, cycle rickshaws and animal drawn carts. This will lead to complex interaction between the vehicles and study of such traffic behavior needs special attention. The traffic plying on roads in western countries is of characteristics of different vehicles with marginal variation contrary to large variation on Indian roads. This will result in increased interactions between vehicles; then they tend to move in clusters rather than one after the other. Further two or three wheelers such as scooters, cycles, and cycle rickshaws contribute to this because of their easy maneuverability.

The traffic on Indian roads consists of bi-directional freedom traffic such as two or three wheeled vehicles and uni-directional vehicles such as four wheelers. While the above tend to overtake or turning or crossing or turn right even if a small gap is available. Hence, to determine the intersection capacity traffic engineer requires a clear understanding of gaps being accepted or rejected by various modes of traffic.

Besides, in these mixed traffic conditions, users do not usually follow lane discipline and can occupy any lateral position on the road. Under these conditions, capacity of an unsignalized intersection is difficult to be determined and becomes a very interesting field of highway capacity study.
There are several types of capacity analysis models for unsignalized intersections such as empirical model based on regression technique and gap acceptance model based on probability theory. The third approach is the conflict technique which was based on the mathematical formulation of interaction and impact between flows at an intersection.

1.1 TYPES OF INTERSECTION CONTROL

As discussed earlier, with increase in complexity of movement, the quantification of the quality of traffic at any intersection is inevitable at every stage of its planning and designing.Based on controls adopted, intersections can be classified as

- Stop and Yield sign control Intersections
- Signalized Intersections

ANNAMACHARYA INSTITUTE OF TECHNOLOGY & SCIENCES

Piglipur, Batasingaram, Hayathnagar (M), Hyderabad, R.R.Dist-501512

3.3.2 Number of Research papers per teachers in the journals notified on UGC website during the last five years.

YEAR	2018-19	2017-18	2016-17	2015-16	2014-15
NO.OF.PAPERS PUBLISHED	0	08	04	0	1

Title of paper	Name of the author/s	Department of the teacher	Name of journal	Year of publication	ISSN number
2017-2018					
An Innovative Non-Intrusive Road Driver Assistance System for Vital Signal Monitoring	Kottapally Chandana K.Ashok Kumar	ECE	International Journal of Research	Oct-17	e-ISSN: 2348-6848 p-ISSN: 2348-795X
Analysis of Data Hiding Techniques in Encrypted Images – A Survey	Seelam Ramakanth Reddy K. Ashok Kumar	ECE	International Journal of Research	Nov-17	e-ISSN: 2348-6848 p-ISSN: 2348-795X
A Wireiess IoT System towards Gait Detect in Stroke Patients	Sajjala Indu K. Ashok Kumar	ECE	International Journal of Research	Oct-17	e-ISSN: 2348-6848 p-ISSN: 2348-795X
Designing Of Navigation System for Blind People Using GPS & GSM Techniques	Nathi Ramyalatha K. Ashok Kumar	ECE	International Journal of Research	Nov-17	e-ISSN: 2348-6848 p-ISSN: 2348-795X
Portable Roadside Sensors For Vehicle Counting, Classification And Speed Measurement	K Ramesh P. Rajeshwar	ECE	International Journal of Research	Oct-17	e-ISSN: 2348-6848 p-ISSN: 2348-795X
A proposed Biometric Multi-server Confirmation Protocol by RFID Elegant cards	B Sravan Kumar P. Rajeshwar	ECE	International Journal of Research	Oct-17	e-ISSN: 2348-6848 p-ISSN: 2348-795X
Microcontroller Base Normal Engine Lock Scheme Intended For Drunken Drivers	Bhukya Shivanaresh R V Prasad Bhookya	ECE	International Journal of Research	Oct-17	e-ISSN: 2348-6848 p-ISSN: 2348-795X

A Sensor Based Device To Monitor The Kitchen	G.Praveena	ECE	International Journal of Research	Oct-17	e-ISSN: 2348-6848 p-ISSN: 2348-795X
			2016-2017		
Implimentation of high-tech agriculture solar fense security with soil humidity based automatic irrigation system and voice alert on PIR live human detection	M.Priyanka G.Praveena	ECE	International Journal Of magazine of engineering, technology , management and resources	Aug-16	ISSN: 2348-4845
School Children Transportation And Safety Enhancement System Based On RFID	K. Ashok Kumar Nitesh Gaikwad	ECE	International journal of advanced Technology and Innovative research	Oct-16	ISSN: 2348-2370
Iterative Receiver for Flip-OFDM in Optical Wireless Communication	K. Vijaya Sheshi rekha	ECE	International journal of Innovative Technologies	Sep-16	ISSN: 2321-8665
IOT Based Secured Smart Home Automation Using Raspberry Pi	R V Prasad Bhookya Nitesh Gaikwad	ECE	International Journal Of magazine of engineering, technology , management and resources	Mar-17	ISSN: 2348-4748
			2014-2015		
A NOVEL MODIFIED HYBRID ALGORITHM TO REDUCE PAPR IN OFDM	D. Sandhya; Nitesh Gaikwad	ECE	INTERNATIONAL JOURNAL OF PROFESSIONAL ENGINEERING STUDIES	Jul-17	Volume II/Issue 4/JULY 2014

PRINCIPAL
Annamacharya Institute of
Technology & Sciences
Piglipur (V), Batasingaram (Post),
Abdullapurmet (M), R.R.Dist.HYD-501 512.

2017-18

International Journal of Research
Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 13
October 2017

An Innovative Non-Intrusive Road Driver Assistance System for Vital Signal Monitoring

[1]Kottapally Chandana
MAIL: kottapally.chandana@gmail.com
Master of Technology
Annamacharya Institute of Technology and Sciences, Blatasingaram, Hayat Nagar, Rangareddy District, Hyderabad, Telangana-500075.

[2]Mr. K.ASHOK KUMAR
Email id: akkonduru@gmail.com
Associate. Professor, M.Tech.
Annamacharya Institute of Technology and Sciences, Blatasingaram, Hayat Nagar, Rangareddy District, Hyderabad, Telangana-500075.

Abstract:

This paper describes an in-vehicle nonintrusive bio potential measurement system for driver health monitoring and fatigue detection. Previous research has found that the physiological signals including eye features, electrocardiography (ECG), electroencephalography (EEG) and their secondary parameters such as heart rate and HR variability are good indicators of health state as well as driver fatigue. A conventional bio potential measurement system requires the electrodes to be in contact with human body. This not only interferes with the driver operation, but also is not feasible for long-term monitoring purpose. The driver assistance system in this paper can remotely detect the bio potential signals with no physical contact with human skin. With delicate sensor and electronic design, ECG, EEG, and eye blinking can be measured. Experiments were conducted on a high fidelity driving simulator to validate the system

monitor the health state of drivers. For medical-assistance systems, the reliable measurement of vital signals such as electroencephalography (EEG) and electrocardiography (ECG) is one of the most important features [1]. EEG, the recording of electrical activity along the scalp, reflects the brain activities and is widely used in the diagnosis of coma and encephalopathy. ECG and the secondary parameters including heart rate (HR) and heart rate variability (HRV) are key indicators of the cardiac health state. The stressful condition of driving and the possible sudden scenarios on the road, e.g., fatal traffic accidents, may cause severe effects especially on the drivers with chronic diseases [2]. Therefore, a driver assistance system that can monitor the multiple vital signals during driving is highly desirable for elderly drivers or drivers with chronic diseases.

For drivers at all ages, drowsiness is one of the most prevalent root causes of accidents. It leads to nearly 17% of all fatal crashes in recent years based

International Journal of Research
Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 14
November 2017

Analysis of Data Hiding Techniques in Encrypted Images – A Survey

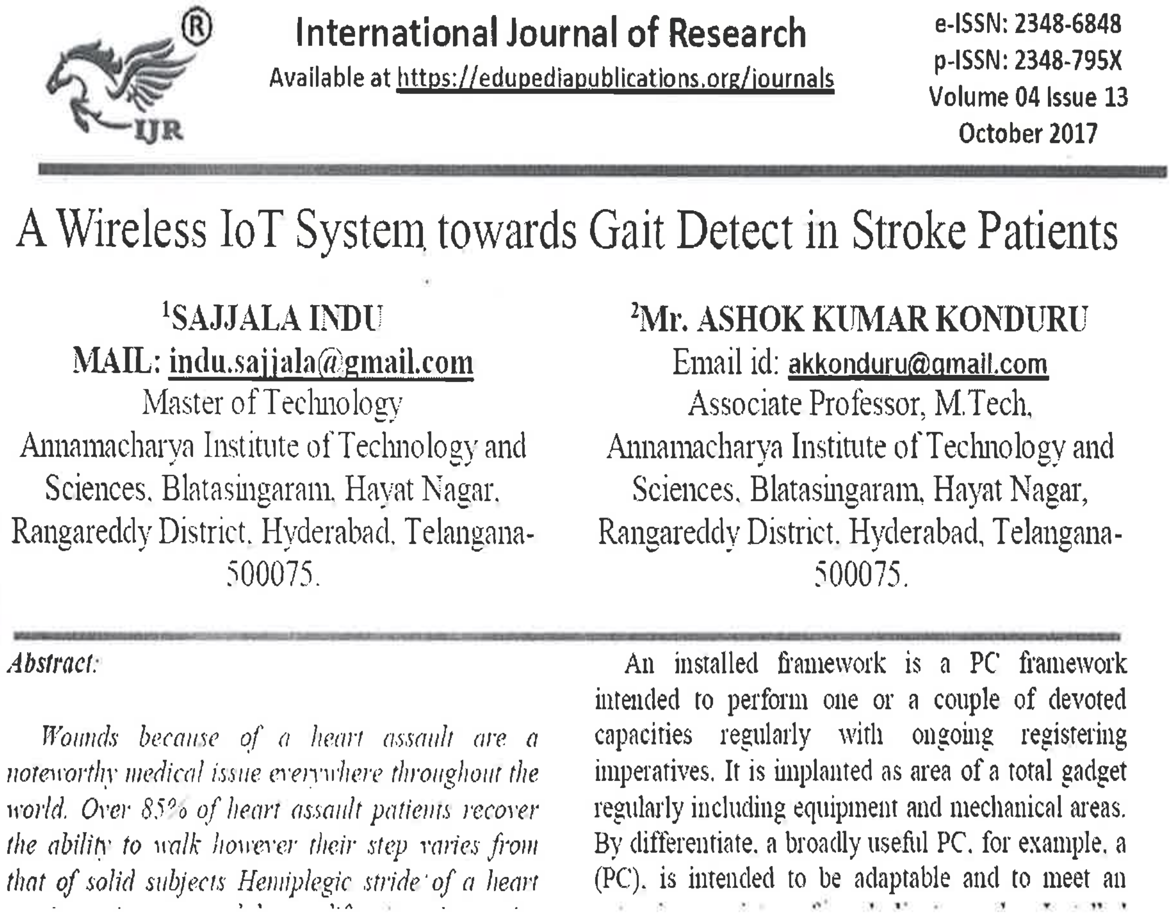
[1]Mr. SEELAM RAMAKANTH REDDY,[2]Mr. ASHOK KUMAR KONDURU

1.Pg Scholar, Department Of ECE, Annamacharya Institute Of Technology And Sciences,Piglipur, Batasingaram(V), Hayathnagar(M), Ranga Reddy(D),Hyderabad.

2. Assoc.Professor and Head of the Department, Department Of ECE, Annamacharya Institute Of Technology And Sciences,Piglipur, Batasingaram(V), Hayathnagar(M), Ranga Reddy(D),Hyderabad

ABSTRACT:

As the use digital techniques for transmitting and storing images are increasing; it is becoming an important issue how to protect the confidentiality, integrity and authenticity of images. This study paper extends to the copying of the data need to be restricted. However encryption does not provide overall protection. Once the encrypted data are decrypted, they can be freely distributed or manipulated. This problem can be solved by hiding some ownership data into the



International Journal of Research
Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 13
October 2017

A Wireless IoT System towards Gait Detect in Stroke Patients

[1]SAJJALA INDU
MAIL: indu.sajjala@gmail.com
Master of Technology
Annamacharya Institute of Technology and Sciences, Blatasingaram, Hayat Nagar, Rangareddy District, Hyderabad, Telangana-500075.

[2]Mr. ASHOK KUMAR KONDURU
Email id: akkonduru@gmail.com
Associate Professor, M.Tech,
Annamacharya Institute of Technology and Sciences, Blatasingaram, Hayat Nagar, Rangareddy District, Hyderabad, Telangana-500075.

Abstract:

Wounds because of a heart assault are a noteworthy medical issue everywhere throughout the world. Over 85% of heart assault patients recover the ability to walk however their step varies from that of solid subjects Hemiplegic stride' of a heart

An installed framework is a PC framework intended to perform one or a couple of devoted capacities regularly with ongoing registering imperatives. It is implanted as area of a total gadget regularly including equipment and mechanical areas. By differentiate, a broadly useful PC, for example, a (PC), is intended to be adaptable and to meet an

International Journal of Research
Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 14
November 2017

Designing Of Navigation System for Blind People Using Gps & Gsm Techniques

[1]NATHI RAMYALATHA,[2]ASHOK KUMAR KONDURU

1.Pg Scholar, Department Of ECE, Annamacharya Institute Of Technology And Sciences,Piglipur, Batasingaram(V), Hayathnagar(M), Ranga Reddy(D),Hyderabad

2. Assoc.Professor and Head of the Department, Department Of ECE, Annamacharya Institute Of Technology And Sciences,Piglipur, Batasingaram(V), Hayathnagar(M), Ranga Reddy(D),Hyderabad

ABSTRACT: In this paper the current state of research and development on global positioning systems (GPS)-based navigation systems for the visually impaired. In this project a walking stick has been designed to help the blind person to detect obstacles and navigate towards the destination. The proposed walking stick consists of a microcontroller, infrared sensors, a GPS receiver, label surface detection, a buzzer and a vibrating motor. The detection of obstacles is done by an array of infrared sensors. The GPS receiver has been used for navigation purpose as

International Journal of Research
Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 13
October 2017

Portable Roadside Sensors for Vehicle Counting, Classification, And Speed Measurement

[1] Kolukulapally Ramesh
Mail: rameshkolukulapally94@gmail.com
Master of Technology
Annamacharya Institute of Technology and Sciences, Blatasingaram, Hayat Nagar, Rangareddy District, Hyderabad, Telangana-500075.

[2]Mr. P. Rajeshwar
Email: rajpakala05@gmail.com
Asst. Professor, M.Tech,
Annamacharya Institute of Technology and Sciences, Blatasingaram, Hayat Nagar, Rangareddy District, Hyderabad, Telangana-500075.

Abstract:

This Project focuses on the development of a portable roadside magnetic sensor system for vehicle counting, classification, and speed measurement. The earliest magnetic field detectors allowed navigation over trackless oceans by sensing the earth's magnetic poles. Magnetic field sensing has vastly expanded as industry has adapted the requirement for superior robots made speedier, more exact and smart robots utilizing new robots control gadgets, new drives and propelled control calculations. This venture depicts another efficient arrangement of robot control frameworks. The introduced robot control framework can be utilized for various refined robot applications.

International Journal of Research
Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 13
October 2017

A Proposed Biometrics-Based Multi-Server Confirmation Protocol by RFID Elegant Cards

[1]Birudu Sravan Kumar
Mail Id: sravan.411@gmail.com
Master of Technology
Annamacharya Institute of Technology and Sciences, Blatasingaram, Hayat Nagar, Rangareddy District, Hyderabad, Telangana-500075.

[2]Mr. P. Rajeshwar
Email: rajpakala05@gmail.com
Asst. Professor, M.Tech,
Annamacharya Institute of Technology and Sciences, Blatasingaram, Hayat Nagar, Rangareddy District, Hyderabad, Telangana-500075.

Abstract:

Recently, in 2014, He and Wang proposed a robust and green multi-server authentication scheme using biometrics-based clever card and elliptic curve cryptography (ECC). In this paper, we first analyze He-Wang's scheme and display that their scheme is

keep them .But the traditional security device is not presenting the higher protection due to the fact in traditional security gadget a person can open the lockers using keys. Sometimes the keys will be stolen. Then the person will practice for new keys but the time period is longer to get new keys so as opposed to the usage of this protection gadget I

International Journal of Research
Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 13
October 2017

Microcontroller Base Normal Engine Lock Scheme Intended For Drunken Drivers

[1]Bhukya Shivanaresh
Mail Id: Shivanareshb@Gmail.Com
Master Of Technology (M.Tech Student)
Annamacharya Institute of Technology and Sciences. Piglipur. Blatasingaram. Hayat Nagar, Ranga Reddy District. Hyderabad. Telangana. 501512.

[2]Mr. R V Prasad Bhookya
Email: Prasad.07461@gmail.com
Asst. Professor. M.Tech
Annamacharya Institute of Technology and Sciences. Piglipur. Blatasingaram, Hayat Nagar, Ranga Reddy District, Hyderabad, Telangana, 501512.

Abstract:

Most of nowadays, we pay attention lot of injuries due to drunken using. Drunken drivers will now not be in solid circumstance and so the rash using is the his cellular smartphone. This gadget is composed of a vibration sensor, Microcontroller and a GSM Modem. The Microcontroller approaches this records and this processed data is despatched to the person/proprietor the usage of GSM modem.

International Journal of Research
Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 13
October 2017

A Sensor-Based Device To Monitor The Kitchen

Shaik Kamlapuram Mujeebullah & G Praveena
1. Pgscholar. Department of Ece. Annamacharya Institute of Technology and Science. Hyderabad
2. Asst. Professor. Department of Ece. Annamacharya Institute of Technology and Science. Hyderabad

ABSTRACT:

Using the advancements in Internet technologies and Wireless Sensor Systems (WSN), a brand new trend within the era of ubiquity has been recognized. Kitchen atmosphere monitoring is among the important measures to become carefully supervised in tangible-here we are at safety, security and luxury of individuals. This technique finds a large application in places that physical presence isn't feasible constantly. The ZigBee tool and ARM1176JZF-S microcontroller are kitchen atmosphere instantly continues to be reported within this paper. The machine can monitor the status of kitchen and send an e-mail and/or perhaps an alert SMS via GSM network instantly, when the conditions get abnormal, to some concerned government bodies cell phone. Customers can monitor and control transducers on active.

Keywords: GSM, Microcontroller, Remote Monitoring, Sensor, ZigBee.

2016-17

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal
IJMETMR

Implementation of Hi-Tech Agricultural Solar Fence Security with Soil Humidity Based Automatic Irrigation System and Voice Alert on PIR Live Human Detection

Markala Priyanka
M.Tech, Embedded Systems,
Annamacharya Institute of
Technology and Sciences.

Mr.Ashok Kumar Konduru
Associate Professor & HOD,
Department of ECE,
Annamacharya Institute of
Technology and Sciences.

Mrs.G.Praveena
Assistant Professor,
Annamacharya Institute of
Technology and Sciences.

Abstract:

Irrigation system in India has given a high priority in economic development. Many new concepts are being developed to allow agricultural automation to flourish and deliver its full potential. To take full advantage of these technologies, we should not just consider the implication of developing a new single technology but should look at the wider issues for complete development of a system. Implementation of Hi-tech Agricultural Solar Fence Security with soil Humidity Based Automatic irrigation system and voice alert on Whenever the dry condition is detected then the motor goes to on condition. Level Sensor is used to indicate the level of water. If water level is LOW or HIGH it will give the buzzer indication. Here we are utilizing solar energy to charge the battery.

IJATIR
International Journal of
Advanced Technology and Innovative Research
www.ijatir.org

ISSN 2348–2370
Vol.08,Issue.14,
October-2016,
Pages:2657-2664

School Children Transportation and Safety Enhancement System Based On RFID

DOREMONI MALLESH KUMAR[1], ASHOK KUMAR KONDURU[2], NITESH GAIKWAD[3]

[1]PG Scholar, Dept of ECE(ES), Annamacharya Institute of Technology and Sciences, Hyderabad, TS, India,
E-mail: dmalleshkumar13@gamil.com.

[2]Assoc Prof & HOD, Dept of ECE, Annamacharya Institute of Technology and Sciences, Hyderabad, TS, India,
E-mail: akkonduru@gmail.com.

[3]Assistant Professor, Dept of ECE, Annamacharya Institute of Technology and Sciences, Hyderabad, TS, India,
E-mail: Niteshgaikwad78@gmail.com.

Abstract: The aim of the project is to design a transportation safety system for school children based on RFID technology. The existing technology over school transportation and child safety system do not exercise any advance technological in electronic devices that may acknowledge the child parent about the arrival of their child to school, the parents are unaware about the information whether their child has attended the school or not, so to eliminate this problem , we system does several tasks, including identifying personal information (Eg. Name) of each student using RFID tag, which will exchange the data with the RFID reader via radio waves and displaying each student name into LCD display. This will let the driver to know the number of students inside the bus and the students who departed from the bus. Moreover, the system has an emergency system that will alert in case if there is a child inside the bus after

ijiTECH International Journal of Innovative Technologies

WWW.IJITECH.ORG

ISSN 2321-8665
Vol.04,Issue.12,
September-2016,
Pages:2310-2313

Iterative Receiver for Flip-OFDM in Optical Wireless Communication

K. VIJAYA[1], ASHOK KUMAR KONDURU[2], SHESHI REKHA[3]

[1]PG Scholar, Dept of ECE, Annamacharya Institute of Technology and Sciences, Hyderabad, TS, India.
E-mail: vijayakota43@gmail.com.

[2]Associate Professor & HOD, Dept of ECE, Annamacharya Institute of Technology and Sciences, Hyderabad, TS, India.
E-mail: akkonduru@gmail.com.

[3]Assistant Professor, Dept of ECE, Annamacharya Institute of Technology and Sciences, Hyderabad, TS, India.
E-mail: Sheshi24@gmail.com.

Abstract: With the rapidly growing demand for data in wireless communications and the significant increase of the number of users, the radio frequency (RF) spectrum become one of the scarcest resources in the world. Motivated by the more and more crowed RF spectrum, optical wireless communications (OWC) has been identified as a promising candidate to complement conventional RF communication, especially for indoor short and medium range data transmission. In the proposed method flip OFDM is used to

IR and visible light provide higher security than RF since both types of signals do not penetrate through walls, and do not get interfered from other rooms or buildings. Both can be used in areas where RF communications is restricted, such as hospitals and airplanes. Nevertheless, OWC still possesses some drawbacks. One restriction is that the available optical transmit power is limited by eye safety standards. In indoor OWC, transmitted data signals can be degraded by multipath propagation, causing OWC channels to be frequency selective

IJEEE

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 4, Issue 3, March 2017)

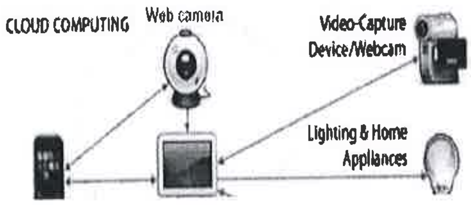
IOT Based Secured Smart Home Automation Using Raspberry Pi

R V Prasad Bhookya
Assistant Professor
Dept. of ECE
Annamacharya Institute of Technology and Sciences
Hyderabad, TS, India
Prasad.07461@gmail.com

Nitesh Gaikwad
Assistant Professor
Dept. of ECE
Annamacharya Institute of Technology and Sciences
Hyderabad, TS, India
niteshgaikwad78@gmail.com

Abstract: This paper deals with the design and implementation of Raspberry pi based IOT concept it means internet of things. In this present generation everything is going on internet itself. So in this project we concentrate totally on the present generation life how they can get security to their home or office and control the devices by using android app just by using internet in there smart phones. The main security is provided by camera module which captures the images and uploads into the internet and also stores the same images in Raspberry pi module SD card. Raspberry pi acts like a small minicomputer it is totally a Linux platform. By just connecting mouse and keyboard we can operate it as minicomputer where we can play games, play videos etc just like our personal laptop work. And also the WI-FI module is used in this project to control the devices from remote location

In terms of complexity embedded systems can range from very simple with a single microcontroller chip, to very complex with multiple units, peripherals and networks mounted inside a large chassis or enclosure.



INTERNATIONAL JOURNAL
OF PROFESSIONAL ENGINEERING STUDIES

Volume II/Issue 4/JULY 2014

A NOVEL MODIFIED HYBRID ALGORITHM TO REDUCE PAPR IN OFDM

D. Sandhya (sandyachakravarthy@gmail.com) [1]

Nitesh Gaikwad (M.tech) (niteshgaikwad78@gmail.com) [2]

[1]Department of ECE, Annamacharya Institute Of Technology And Sciences.

[2]Associate Professor, Department of ECE, Annamacharya Institute Of Technology And Sciences.

ABSTRACT:

Orthogonal Frequency Division Multiplexing (OFDM) is an efficient method of data transmission for high speed communication systems. However, the most disadvantage of OFDM system is that the high Peak to Average Power Ratio (PAPR) of the ,transmitted signals. OFDM carries with it sizable amount of number of independent subcarriers, as a results of that the amplitude of

opposite hand, in FDM system, the carriers ar spaced apart with guard bands in such how that guard bands ar introduced between the different carriers within the frequency domain, which ends in lowering spectrum potency [2].

To reduce the PAPR, many techniques are fictitious [3] that primarily are often divided in 3 classes. First of all, signal distortion techniques that scale back the height

ANNAMACHARYA INSTITUTE OF TECHNOLOGY & SCIENCE

Piglipur, Batasingaram, Hayathnagar (M), Hyderabad, R.R.Dist-501512

3.3.2 Number of Research papers per teachers in the journals notified on UGC website during the last five years.

YEAR	2018-19	2017-18	2016-17	2015-16	2014-15
NO.OF.PAPERS PUBLISHED	2	2	6	02	2

Title of paper	Name of the author/s	Department of the teacher	Name of journal	Year of publication	ISSN number
			2018-2019		
A STUDY ON POWER QUALITY AND RELIABILITY COMPREHENSIVE POWER DISPENSATION OF SMALL SCALE POWER SYSTEM AND SUBSEQUENT MICRO-GRIDS	U.NARENDER	EEE	JASC	MAY-2019	ISSN: 1076-5131
ACTIVE POWER QUALITY IMPROVEMENT AND DISTRIBUTED CONTROL IN HYBRID AC/DC MICROGRIDS	U.NARENDER	EEE	International Journal of Research	MAY-2019	ISSN: 2236-6124
			2017-2018		
An interleaved High power flyback inverter for photovoltaic applications	G.BHASKAR RAO / U.NARENDER	EEE	IJSETR	FEB-2017	ISSN: 2319-8885
A novel soft switching DC - DC Converter with distributed energy storage voltage capability	U.NARENDER	EEE	IJSETR		ISSN: 2319-8885
			2016-2017		
"A Novel Solar Power Optimizer Implementation for Power Quality Improvement of DC Distribution System	J.SHANKAR	EEE	IJITECH	JULY-2017	ISSN: 2321-8665
An interleaved High power flyback inverter for photovoltaic applications	U.NARENDER	EEE	IJSETR	FEB-2017	ISSN:2319-8885

OPERATION AND CONTROL OF AN IMPROVED PERFORMANCE INTERACTIVE DSTATCOM	J. SREEDHAR	EEE	IJATIR	DEC-2016	ISSN:2348-2370
HIGH STEP-UP DUAL SWITCH CONVERTER WITH COUPLED INDUCTOR AND VOLTAGE MULTIPLYER FOR GRID CONNECTED SYSYTEM	J. SREEDHAR	EEE	IJIEMR	JULY-2017	ISSN:2456-5083
DESIGN AND PERFORMANCE OF VOLTAGE CONTROL DSTATCOM BASED IMPROVEMENT OF POWER QUALITY IN DISTRIBUTION SYSTEM	M.CHANDRA SHEKHAR	EEE	IJERAD	DEC-2017	ISSN: 2017-0312
POWER QUALITY IMPROVEMENT BY VOLTAGE CONTROL USING DSTATCOM	M.CHANDRA SHEKHAR	EEE	IJR	NOV-2017	ISSN: 2348-6848
2015-2016					
POWER QUALITY IMPROVEMNET FOR MICRO GRID WITH MULTIPLE ENERGY SOURCE	M . CHANDRASHEKHAR	EEE	IJERSTEE	MARCH-2015	ISSN: 2319-7463
PERFORMANCE IMPROVEMENT OF RENEWABLE ENERGY SYSTEM THROUGH ACTIVE POWER FILTER FOR MEETING THE ENERGY DEMAND AND POWER QUALITY IMPROVEMENT	M.B.HEMANTH	EEE	IJAREEIE	JUNE-2016	ISSN: 2320-3765
2014-2015					
New Hybrid Power Conditioner for Suppressing Harmonics and Neutral Line Current in Three Phase Four Wire Distribution Power System	J.SHANKAR	EEE	JSETR	SEP-2015	ISSN: 2319-8885
A New Bidirectional Intelligent Semiconductor Transformer for Smart Grid Applications	J.SHANKAR	EEE	IJISET	AUG-2015	ISSN: 2348-7968

A STUDY ON POWER QUALITY AND RELIABILITY COMPREHENSIVE POWER DISPENSATION OF SMALL SCALE POWER SYSTEM AND SUBSEQUENT MICRO-GRIDS

Mr.U.NARENDER
Research Scholar,
Department of Electrical Engineering,
Shri JJT University, Rajasthan, INDIA.
narender0866@yahoo.co.in

Dr. Mrs. Anupama A. Deshpande
Ex-Principal, Ex-Head of Electrical Dept.
& Ex- IEDC Coordinator
Atharva College of Engineering (ACE)
Mumbai, Maharashtra, INDIA
mangala.d.2000@gmail.com

ABSTRACT:

The significance of an adaptable AC dispersion framework gadget for micro-grid applications. The device intends to improve the power quality and dependability of the general power conveyance framework that the micro-grid is associated with Broadened Kalman channels are additionally contemplated for recurrence following and to extricate the symphonies spectra of the lattice voltage and the heap flows in the micro-grid. Likewise these paper high lights on DG gathering so as to orchestrate the venture of benefits, the nature of intensity supply and the participation with the current power lattice. The heritage worldview for power administration in the greater part of the jolted present reality depends on the concentrated generation transmission-circulation framework that developed under a controlled domain. In this specific situation, another worldview is developing wherein power generation is personally installed with the heap in micro-grids. Up to now, these units just infuse dynamic power depending from the accessibility of their essential source. In future from one perspective DG units need to add to framework strength, however then again DG units can give extra functionalities so as to offer a surplus incentive for the client. Hence particularly inverter-coupled frameworks are appropriate. Extra usefulness could be improvement of Power Quality and Reliability (PQR), yet additionally crest shaving, arrangement of control vitality or receptive power remuneration is possible. And here given the future development of micro-grid and its features also given.

Key words: Distributed generation, Power system operation, Power system reliability, Electric power quality, Future Micro-grids.

1.0 INTRODUCTION:

For traditional power distribution framework, the idea of micro-grid has offered shoppers an unwavering quality and decrease in absolute vitality misfortunes and it has turned into a promising option. While associating micro-grid to the dissemination matrix, the effect of power quality (PQ) issue on the general power framework execution must be considered. These PQ issue incorporates voltage and recurrence deviation in the lattice voltage and sounds in the voltage and burden flows. To relieve these issue different types of gear, for example, dynamic channels, continuous power supplies, active voltage restorers, and UPQC are generally introduced by the purchasers to secure their heaps and framework against PQ aggravations in dissemination organize. In any case, these gadgets are introduced at the purchaser sides and the PQ issues that they are proficient to deal with are typically

ACTIVE POWER QUALITY IMPROVEMENT AND DISTRIBUTED CONTROL IN HYBRID AC/DC MICROGRIDS

Mr.U.NARENDER[1] , Dr. Mrs. ANUPAMA A. DESHPANDE[2]

[1]Research Scholar, Department of Electrical Engineering, Shri JJT University, Rajasthan, INDIA.
narender0866@yahoo.co.in

[2]Ex-Principal, Ex-Head of Electrical Dept. & Ex- IEDC Coordinator, Atharva College of Engineering (ACE) ,Mumbai, Maharashtra, INDIA
mangala.d.2000@gmail.com

ABSTRACT:

Combining the DC microgrid and the dominated AC system forms the scenario hybrid AC/DC microgrid, which would be, in concept, the presence of both DC and AC microgrids with sources, storages, loads and appropriate interlinking converters (ICs) tied between them. Hybrid AC/DC microgrid has been becoming a popular concept to provide an effective solution for unlimited large-scale integration of various DGs and distributed storages (DSs) because of its higher efficiency and better compatibility. This paper proposes a distributed control strategy that considers several source characteristics to achieve reliable and efficient operation of a hybrid ac/dc micro grid.The hierarchical control scheme for standalone DC microgrids, the fully decentralized control for hybrid AC/DC microgrids, the distributed control for hybrid AC/DC/DS microgrids and power quality improvement for hybrid AC/DC microgrid have been verified. The proposed hybrid ac/dc microgrid is composed of converters and distributed generation units that include renewable energy sources (RESs) and energy storage systems (ESSs). The proposed control strategy is verified in various scenarios experimentally and by simulation.

Keywords: ac/dc hybrid microgrid, power quality,harmonic compensation, reactive power compensation

INTRODUCTION:

Alternative current (AC) has been the dominant power supply medium for over a century since the end of "the war of currents" [10] in which Thomas Edison and George Westinghouse became adversaries due to Edison's promotion of direct current (DC) for electric power distribution over AC advocated by Westinghouse. War of the currents was ultimately won by AC, and has been the platform for electrical transmission across the world since then. The key behind AC's victory was the invention of the transformers which could easily step-up the voltage levels for long distance power transfer with lower transmission losses. The points of AC being the standard choice include easier transformation into different levels for various applications, capability of long distance power transmission and inherent characteristics from the fossil energy driven rotating machine. AC power system gradually became the top engineering achievement of the 20th century. However, problems along with the development, such as high energy costs, aging of current power system infrastructure and limited funds to construct new large power plants and long distance transmission lines, constraint the meet of the growing energy demands.

On the other side, the advantage of DC transmission was re-recognized accompanied with the progress of advanced power electronics techniques. The major application is power electronics-based high voltage DC (HVDC) transmission, which integrates DC penetration inside AC-dominated transmission networks. Since the past two decades, DC grids have shown resurgence due to the development and deployment of renewable DC power sources and their inherent compatibility for various DC loads in industrial systems [1], commercial buildings [3] and residential complex [9]. Reasons of the gaining popularity for DC grids include better compatibility [5], higher efficiency [6] and robust stability [6]. The shift from AC to DC system facilitates easier control of individual load performance, increased integration of renewable energy sources (RESs) and distributed energy storages [10]. This trend calls for a re-examination of the traditional AC power system structure and its efficiency. An alternative solution might be the hybrid AC/DC power system,

International Journal of
Scientific Engineering and Technology Research

www.ijsetr.com

ISSN 2319-8885
Vol.06,Issue.04
February-2017,
Pages:0666-0670

An Interleaved High-Power Flyback Inverter for Photovoltaic Applications

U. NARENDER[1], K. SWAROOPA[2], JADAPALLI SREEDHAR[3], G BHASKAR RAO[4]

[1]Assistant Professor, Dept of EEE, Annamacharya Institute of Technology & Sciences, Hyderabad, TS, India,
E-mail: narender0866@yahoo.co.in

[2]PG Scholar, Dept of EEE, Annamacharya Institute of Technology & Sciences, Hyderabad, TS, India,
E-mail: swaroopa.7102@gmail.com.

[3]Associate Professor & HOD, Dept of EEE, Annamacharya Institute of Technology & Sciences, Hyderabad, TS, India,
E-mail: sreedharmtech@gmail.com.

[4]Assistant Professor, Dept of EEE, Annamacharya Institute of Technology & Sciences, Hyderabad, TS, India,
E-mail: bhaskar.bhaskarg@gmail.com.

Abstract: Solar energy is the most important renewable and freely available source of energy. It has a main advantage of pollution free source. Hence, this source of energy will be the center of vision in the sector of energy market in near future. Fly back inverter has advantages such as simple control loop, electric isolation, and high step up ratio. Therefore, it is an attractive solution for photovoltaic application. This paper is dealt with how these two advantageous concepts are coupled together to give a low ripple and high Total Harmonic Distortion(THD).An advanced interleaved concept is also used in this paper. At the same time, harvesting maximum of solar energy is best in this method as there is a maximum power point tracker (MPPT) for each PV cell. Thus an improved interleaved high power fly back inverter will be efficient for PV application with low ripples. Then, the design is verified and optimized for the best performance based on the simulation results. Finally, a prototype at rated power is built and evaluated under the realistic conditions. The efficiency of the inverter, the total harmonic distortion of the grid current, and the power factor are measured as 90.16%, 4.42%, and 0.998, respectively. Consequently, it is demonstrated that the performance of the proposed system is comparable to the commercial isolated PV inverters in the market, but it may have some cost advantage.

Keywords: Fly Back Converter, Harmonics, Interleaved Converters, Photovoltaic (PV) Inverters.

I. INTRODUCTION

Solar power is the best available renewable energy that has wide scope for harvesting huge amount of electricity. It can be said that an average home has more than enough area in the roof to produce sufficient amount of solar electricity to meet all its requirements. Using an Inverter, the DC power from the PV array can be conveniently converted to AC similar to connection with normal power grid. There are only two primary disadvantages to using solar power: amount of sunlight and cost of equipment. The best way of lowering the cost of solar energy is to improve the cell's efficiency. In this project we have proposed an Improved fly back inverter which operates in BCM mode so as to increase the efficiency. In this paper, the efficiency of inverter can be improved by using fly back topology the topology of the fly back inverter, which consists of three MOSFETs, two diodes, and a fly back transformer. The two outputs from the transformer are connected to the grid, through a common filter circuit, which can switch reciprocally and synchronously with the polarity of the grid voltage. It is simple, has less component count hence reduced cost and has advantage of isolation through transformer. A Pulse Width Modulation (PWM) control circuit to control the duty ratio of the switch. In fly-back circuits, for closed loop output voltage regulation, one needs to feed output voltage magnitude to the PWM controller. Here the Boundary Conduction Mode is used for the purpose of choosing broader frequency range. The energy from sun is transformed into direct current electricity. Maximum power point is a unique operating point supplying maximum power to the load which is present in a PV array. Tracking the maximum power point of the PV array is done to improve the efficiency of the photovoltaic energy system MPPT is an electronic system that operates the Photovoltaic (PV) modules in a manner that allows the modules to produce all the power capable of PV module. MPPT is not a mechanical tracking system that "physically moves" the modules to make them point more directly at the sun.

II. TYPES OF INVERTERS

The basic function of the inverter in a photovoltaic solar power system array is to convert the DC electricity generated by the solar panels into standard AC power. Any photovoltaic system which supplies power to an load must use an inverter to cover the DC power generated into AC power. There are four basic types of inverters commonly used namely Standalone inverters, Grid Tie inverters ,bimodal inverters,

International Journal of
Scientific Engineering and Technology Research

ISSN 2319-8885
Vol.06,Issue.04
February-2017,
Pages:0661-0665

www.ijsetr.com

A Novel on Soft Switching DC-DC Converter with Distributed Energy Source Voltage Capability

U. NARENDER[1], MADUGULA SRIKANTH[2], JADAPALLI SREEDHAR[3], G BHASKAR RAO[4]
[1]Assistant Professor, Dept of EEE, Annamacharya Institute of Technology & Sciences, Hyderabad, TS, India,
E-mail: narender0866@yahoo.co.in
[2]PG Scholar, Dept of EEE, Annamacharya Institute of Technology & Sciences, Hyderabad, TS, India,
E-mail: 229srikanth@gmail.com.
[3]Associate Professor & HOD, Dept of EEE, Annamacharya Institute of Technology & Sciences, Hyderabad, TS, India,
E-mail: sreedharmtech@gmail.com.
[4]Assistant Professor, Dept of EEE, Annamacharya Institute of Technology & Sciences, Hyderabad, TS, India,
E-mail: bhaskar.bhaskarg@gmail.com.

Abstract: A soft-switching dc/dc converter with high voltage gain is proposed in this paper. It provides a continuous input current and high voltage gain. Moreover, soft-switching characteristic of the proposed converter reduces switching loss of active power switches and raises the conversion efficiency. The reverse-recovery problem of output rectifiers is also alleviated by controlling the current changing rates of diodes with the use of the leakage inductance of a coupled inductor. Hybrid power system consists of a combination of renewable energy sources such as: photovoltaic (PV), wind generators, hydro, etc., to charge batteries and provide power to meet the energy demand. Finally, a simplified design procedure is proposed in hybrid system by using stand-alone application.

Keywords: Active Clamp, DC-DC Converter, High Step Up, Switched Capacitor, Voltage Doubler Circuit.

I. INTRODUCTION

Recently, the demand for dc/dc converters with high voltage gain has increased. The energy shortage and the atmosphere pollution have led to more researches on the renewable and green energy sources such as the solar arrays and the fuel cells. Moreover, the power systems based on battery sources and super capacitors have been increased. Unfortunately, the output voltages of these sources are relatively low. Therefore, the step-up power conversion is required in these systems. Besides the step-up function, the demands such as low current ripple, high efficiency, fast dynamics, light weight, and high power density have also increased for various applications. Input current ripple is an important factor in a high step-up dc/dc converter. Especially in the fuel cell systems, reducing the input current ripple is very important because the large current ripple shortens fuel cell's lifetime as well as decreases performances. Therefore, current fed converters are commonly used due to their ability to reduce the current ripple. In applications that require a voltage step-up function and a continuous input current, a continuous-conduction-mode (CCM) boost converter is often used due to its advantages such as continuous input current and simple structure. However, it has a limited voltage gain due to its parasitic components. The reverse-recovery problem of the output diodes is another important factor in dc/dc converters with high voltage gain.

II. PROPOSED SYSTEM PROJECT DESCRIPTION

A non isolated converter using voltage multiplier cell is introduced. It can achieve high voltage gain without large duty ratio operation. However, to obtain high conversion ratio multiple of such cell has to be used. Switched capacitor or switched inductor based converters are given in literatures. Switched capacitor converter proposed have advantage of lack of magnetic components and good line regulation. For high gain applications, these converters require more number of switched capacitor and switched inductor cells. This increases the component counts.

A. Active clamp

Forward converters with active-clamp reset offer multiple benefits to designers and are presently finding wide use. Power converters based on the forward topology are an excellent choice for applications where high efficiency and good power handling capability is required in the 50 to 500W power range. While the popularity of forward topology is based upon many factors, designers have been primarily drawn to it's simplicity, performance, and efficiency. The forward converter is derived from the buck topology. The main difference between the two topologies is that the transformer employed in the forward topology provides input-output ground isolation as well as a step-down or step-up function. The transformer in a forward topology does not

International Journal of Innovative Technologies
WWW.IJITECH.ORG
ISSN 2321-8665
Vol.05,Issue.07,
July-2017,
Pages:1305-1311

A Novel Solar Power Optimizer Implementation for Power Quality Improvement of DC Distribution System

VEERABHADRESHWAR VEDALA[1], J. SHANKAR[2]
[1]Assistant Professor, Dept of EEE, Vignan Institute of Technology & Science, Hyderabad, TS, India,
Email: eshwar.vedala@gmail.com.
[2]Assistant Professor, Dept of EEE, Annamacharya Institute of Technology & Science, Hyderabad, TS, India,
Email: shankar.jngm@gmail.com.

Abstract: In this paper a new topology with a reversing voltage component is proposed which will improve the multilevel performance by compensating the disadvantages of increased number of components, complex pulse width modulation control method and voltage balancing problem. This topology requires fewer components compared to existing inverter topologies (particularly in higher levels) and requires fewer carrier signals and gate drives. Therefore the overall cost and complexity are greatly reduced particularly for higher output voltage levels. This paper describes the general multilevel inverter schematic and modified circuit having only seven switches for seven- level multilevel inverter using reverse voltage topology. A general method of multilevel modulation phase disposition (PD) SPWM is utilized to drive the inverter and can be extended to any number of voltage levels. The simulation of a modified seven level multilevel inverter using reverse voltage topology is also presented here.

Keywords: Multilevel Inverter, Power Electronics, SPWM, Topology.

I. INTRODUCTION

MLI's are used for high power and high voltage applications. MLI's have unique structure which makes it possible to reach high voltages with less harmonic content. Inverter is a device which converts dc power to ac power. Two level inverters require high switching frequency and disadvantages are less efficiency, high cost and high switching losses. Various PWM strategies are required to get high quality output which leads to high switching losses. MLI's are introduced to overcome these problems. MLI is able to synthesize output voltages with reduced harmonic distortion and lower electromagnetic interference. The advantages of MLI are improvement in staircase waveform quality, less input current distortion, lower electromagnetic interference. MLI's are used in drives, PV systems and automotive applications. The harmonic content of the output voltage waveform decreases as the number of output voltage increases. MLI's are mainly classified as cascaded MLI, diode clamped MLI, flying capacitor MLI. The control method of cascaded H- bridge MLI because it doesn't have any clamping diode and flying capacitor.

Cascaded MLI reaches higher reliability and this is used for large automotive electric drives. The main disadvantage is the increase in number of power switches that normally contributes to the complexity in con-trolling power switches. Many methods have been developed to decrease the number of switches. Modulation strategies applied to MLI's are selective harmonics elimination, carrier based PWM, space vector modulation, and fundamental frequency modulation. The PWM control is the most efficient method of controlling output voltage within the inverters. The carrier based PWM schemes used for MLI's is much more efficient, realized by the intersection of modulating signal with triangular carrier waveform. This paper is based on seven level inverter with reverse voltage topology which requires less number of switches than conventional topologies this paper aims at generation of carrier based PWM scheme using PD method and can control output voltage and frequency and reduce the harmonic components in load currents here PD SPWM use ((n-1)/2) carriers to drive the inverter. In PD, all the carrier waveforms are in phase.

II. MULTILEVEL INVERTER USING REVERSE VOLTAGE TOPOLOGY

Conventional cascaded MLI's require large number of switches and the power semiconductor switches are combined to produce an output in positive and negative polarity. In the new topology, there is no need to utilize all the switches in high frequency. This topology separates output voltage into level generation and polarity generation parts. Level generation part generates levels in positive polarity and polarity generation part generates the polarity of the output voltage. Level generation part needs high frequency switches and polarity generation part requires low frequency switches operating at line frequency. Fig. 1 shows schematic diagram of a single phase seven- level reverse voltage topology. This MLI's can be increased to higher voltage levels by increasing middle section. This topology requires less switches and it can be applied to three phase application. The PD SPWM for proposed topology needs only half the number of conventional carriers for SPWM. PD SPWM for seven level conventional inverters requires six carriers, but in the proposed system only three carriers are need-ed. MLI control

www.ijsetr.com

International Journal of
Scientific Engineering and Technology Research

ISSN 2319-8885
Vol.06,Issue.04
February-2017,
Pages:0666-0670

An Interleaved High-Power Flyback Inverter for Photovoltaic Applications

U. NARENDER[1], K. SWAROOPA[2], JADAPALLI SREEDHAR[3], G BHASKAR RAO[4]

[1]Assistant Professor, Dept of EEE, Annamacharya Institute of Technology & Sciences, Hyderabad, TS, India,
E-mail: narender0866@yahoo.co.in

[2]PG Scholar, Dept of EEE, Annamacharya Institute of Technology & Sciences, Hyderabad, TS, India,
E-mail: swaroopa.7102@gmail.com.

[3]Associate Professor & HOD, Dept of EEE, Annamacharya Institute of Technology & Sciences, Hyderabad, TS, India,
E-mail: sreedharmtech@gmail.com.

[4]Assistant Professor, Dept of EEE, Annamacharya Institute of Technology & Sciences, Hyderabad, TS, India,
E-mail: bhaskar.bhaskarg@gmail.com.

Abstract: Solar energy is the most important renewable and freely available source of energy. It has a main advantage of pollution free source. Hence, this source of energy will be the center of vision in the sector of energy market in near future. Fly back inverter has advantages such as simple control loop, electric isolation, and high step up ratio. Therefore, it is an attractive solution for photovoltaic application. This paper is dealt with how these two advantageous concepts are coupled together to give a low ripple and high Total Harmonic Distortion(THD).An advanced interleaved concept is also used in this paper. At the same time, harvesting maximum of solar energy is best in this method as there is a maximum power point tracker (MPPT) for each PV cell. Thus an improved interleaved high power fly back inverter will be efficient for PV application with low ripples. Then, the design is verified and optimized for the best performance based on the simulation results. Finally, a prototype at rated power is built and evaluated under the realistic conditions. The efficiency of the inverter, the total harmonic distortion of the grid current, and the power factor are measured as 90.16%, 4.42%, and 0.998, respectively. Consequently, it is demonstrated that the performance of the proposed system is comparable to the commercial isolated PV inverters in the market, but it may have some cost advantage.

Keywords: Fly Back Converter, Harmonics, Interleaved Converters, Photovoltaic (PV) Inverters.

I. INTRODUCTION

Solar power is the best available renewable energy that has wide scope for harvesting huge amount of electricity. It can be said that an average home has more than enough area in the roof to produce sufficient amount of solar electricity to meet all its requirements. Using an Inverter, the DC power from the PV array can be conveniently converted to AC similar to connection with normal power grid. There are only two primary disadvantages to using solar power: amount of sunlight and cost of equipment. The best way of lowering the cost of solar energy is to improve the cell's efficiency. In this project we have proposed an Improved fly back inverter which operates in BCM mode so as to increase the efficiency. In this paper, the efficiency of inverter can be improved by using fly back topology the topology of the fly back inverter, which consists of three MOSFETs, two diodes, and a fly back transformer. The two outputs from the transformer are connected to the grid, through a common filter circuit, which can switch reciprocally and synchronously with the polarity of the grid voltage. It is simple, has less component count hence reduced cost and has advantage of isolation through transformer. A Pulse Width Modulation (PWM) control circuit to control the duty ratio of the switch. In fly-back circuits, for closed loop output voltage regulation, one needs to feed output voltage magnitude to the PWM controller. Here the Boundary Conduction Mode is used for the purpose of choosing broader frequency range. The energy from sun is transformed into direct current electricity. Maximum power point is a unique operating point supplying maximum power to the load which is present in a PV array. Tracking the maximum power point of the PV array is done to improve the efficiency of the photovoltaic energy system MPPT is an electronic system that operates the Photovoltaic (PV) modules in a manner that allows the modules to produce all the power capable of PV module. MPPT is not a mechanical tracking system that "physically moves" the modules to make them point more directly at the sun.

II. TYPES OF INVERTERS

The basic function of the inverter in a photovoltaic solar power system array is to convert the DC electricity generated by the solar panels into standard AC power. Any photovoltaic system which supplies power to an load must use an inverter to cover the DC power generated into AC power. There are four basic types of inverters commonly used namely Standalone inverters, Grid Tie inverters ,bimodal inverters,

International Journal of
Advanced Technology and Innovative Research

www.ijatir.org

ISSN 2348–2370
Vol.08,Issue.24,
December-2016,
Pages:4681-4685

Operation and Control of an Improved Performance Interactive DSTATCOM

CH. VIJAY KUMAR[1], M. CHANDRASHEKHAR[2], JADAPALLI SREEDHAR[3]

[1]PG Scholar, Dept of EEE(EPS), Annamacharya Institute of Technology & Science, Hyderabad, TS, India,
E-mail: sumanthjaya@gmail.com.

[2]Assistant Professor, Dept of EEE, Annamacharya Institute of Technology & Science, Hyderabad, TS, India,
E-mail: chandrashekhar207@gmail.com.

[3]Associate Professor & HOD, Dept of EEE, Annamacharya Institute of Technology & Science, Hyderabad, TS, India.

Abstract: In modern power system, the switching devices are generally used in combination with unbalanced reactive loads which produces current related Power Quality (PQ) problems by making source currents distorted and unbalanced. Usually a STATCOM is installed to support electricity networks that have a poor power factor and often poor voltage regulation. A Distribution Static Compensator (DSTATCOM) operating in Current Control Mode (CCM) is used to mitigate current related PQ problems. The main objective is to improve the performance of interactive distribution static compensator for address limitations of conventional Current Control Mode (CCM) and Voltage Control Mode (VCM) operations. In CCM operation, the DSTATCOM supplies reactive and harmonic component of load currents to make source currents balanced, sinusoidal, and in-phase with respective phase load voltages In the previous research, the compensator with three single-phase H-bridge Voltage Source Inverters (VSIs), driven by a single DC storage capacitor was used and passive filter capacitor is also used to provide the path for high frequency components for current. Thus the operation and control of the improved performance interactive DSTATCOM is proposed for the continuous and stable load operation. Using the control algorithm, the range of source voltage within which a DSTATCOM should operate in CCM is computed. The advantage of the proposed is that the algorithm depends upon the supply voltage, maximum and minimum feeder impedance and load current. Outside this range, operational mode of DSTATCOM is transferred to VCM. Losses in feeder and VSI are reduced which improves efficiency of the system. The proposed method is implemented in MATLAB/Simulink and shows that the performance is improved with reduced loss, cost, and power rating VSI as compared to the conventional CCM and VCM DSTATCOM operation.

Keywords: Voltage Control Mode, Current Control Mode, Power Factor.

I. INTRODUCTION

Switching devices in combination with unbalanced reactive loads produce current related power quality (PQ) problems by making source currents distorted and unbalanced.A distribution static compensator (DSTATCOM) operating in current control mode (CCM) is used to mitigate current related PQ problems. In CCM operation, the DSTATCOM supplies reactive and harmonic component of load currents to make source currents balanced, sinusoidal, and in-phase with respective phase load voltages. Generally, faults in power system and energization of larger loads create voltage disturbances like sag and swell. Also, integration of intermittent distributed generation causes voltage fluctuations in the distribution system. These voltage disturbances significantly affect the power transfer from the source to load and degrade the performance of sensitive loads. However, conventional CCM operation of DSTATCOM cannot improve the load voltage. This is major limitation of CCM operation of DSTATCOM which considerably restricts its utilization. A DSTATCOM, when operated in voltage control mode (VCM), is one of the most effective device used for load voltage regulation. In VCM operation, the DSTATCOM regulates load voltage at a constant reference value by supplying appropriate fundamental reactive current into the source.

Therefore, VCM operation of DSTATCOM provides stable and continuous operation of the load. However, conventional VCM operation of DSTATCOM maintains an arbitrary chosen voltage of 1.0 p.u. at the load terminal. For this voltage at load terminal, source exchanges reactive power even at normal operating conditions this continuous reactive power exchange results in more reactive current flow in the voltage source inverter (VSI) as well as feeder. Consequently, losses in the VSI and feeder increase. Therefore, VCM operation of DSTATCOM is not required during normal supply conditions. Aforementioned analysis brings the fact that the conventional CCM and VCM operations of DSTATCOM are not required during voltage disturbances and normal disturbances, respectively. This greatly limits utilization of the DSTATCOM. Moreover, recent advancements in device topologies and control algorithms have encouraged customers to look for devices which can provide various operational characteristics with less number of components, reduced cost, weight, and space This paper proposes operation and control of an improved performance interactive DSTATCOM for continuous and stable load operation while addressing

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.o

HIGH STEP-UP DUAL SWITCH CONVERTER WITH COUPLED INDUCTOR AND VOLTAGE MULTIPLIER FOR GRID CONNECTED SYSTEM

[1]G.RADHAKRISHNA, [2]M.CHANDRA SHEKHAR, [3]MR. JADAPALLI SREEDHAR

[1]PG Scholar, Dept of EEE, Annamacharya Institute of Technology & Sciences, Hyderabad, RR (Dt), Telangana, India.

[2]Assistant Professor, Dept of EEE, Annamacharya Institute of Technology & Sciences, Hyderabad, RR (Dt), Telangana, India.

[3]HOD, Dept of EEE, Annamacharya Institute of Technology & Sciences, Hyderabad, RR (Dt); Telangana, India.

ABSTRACT:

A novel high step-up converter, which is suitable for a PV cell is proposed in this paper. The proposed converter is composed of the dual switches structure, the coupled inductor, and voltage multiplier cells in order to achieve the high step-up voltage gain. The dual switches structure is beneficial to reduce the voltage stress and current stress of the switch. In addition, two multiplier capacitors are, respectively, charged during the switch-on period and switch-off period, which increases the voltage conversion gain. Meanwhile, the energy stored in the leakage inductor is recycled with the use of clamped capacitors. Thus, two main power switches with low on-resistance and low current stress are available. As the leakage inductor, diode reverse-recovery problem is also alleviated. Therefore, the efficiency is improved. This paper illustrates the operation principle of the proposed converter; discusses the effect of the leakage inductor; analyzes the influence of parasitic parameters on the voltage gain and efficiency, the voltage stresses and current stresses of power devices; and a comparison between the performance of the proposed converter and the previous high step-up converters is performed. In the ordinary voltage step-up situation, the conventional step-up converters, such as the boost converter can satisfy the voltage step-up requirement. However, in the high step-up situation, the conventional converter cannot achieve a high step-up conversion with high efficiency by extreme duty cycle or high turns ratio because of the parasitic parameters or leakage inductance.

Key Words: Dual switches, high step-up converter, switched capacitor, three-winding coupled inductor, Photo Voltaic System.

I. INTRODUCTION

In recent years, the boost dc/dc converters have been widely used to step up the renewable energy sources in various industrial applications such as ESS, UPS, and EV etc. In those applications, boost dc/dc converters generally step up the voltage to the high voltage output. For that reason, to obtain a high voltage gain, many converter topologies were reported [3]-[6] for this application. Direct

voltage step up using high frequency transformer is a Simple and easily controllable converter providing high gain. Isolated current

International Journal of Engineering Research and Advanced Development
Volume 3, Issue 12, DIP: 18.03.12/20170312
http://www.ijerad.redmac.in |December 2017

Original Research Paper

Design and Performance of Voltage-Controlled DSTATCOM Based Improvement of Power Quality in Distribution System

M.Chandra Shekhar[1]*, P.Renuka[2], Mr.Jadapalli Sreedhar[3]

ABSTRACT

Now a day's most popular devices are facts device because of more flexibility and reliability to remove the power quality issues. In this paper, Due to increasing complexity in the power system, voltage sag is becoming one of the most significant power quality problems in the distribution system. Voltage sag is a short reduction voltage from nominal voltage, occurs in a short time. If the voltage sags exceed two to three cycles, then manufacturing systems making use of sensitive electronic equipments are likely to be affected leading to major problems. It ultimately leads to wastage of resources (both material and human) as well as financial losses. This project tends look at the solving the sag problems and harmonics by using custom power devices such as Distribution Static compensator (D-STATCOM). Proposed scheme follows a new algorithm to generate reference voltage for a distribution static compensator (DSTATCOM) operating in voltage-control mode. The proposed scheme ensures that unity power factor (UPF) is achieved at the load terminal during nominal operation, which is not possible in the traditional method in distribution system. Also, the compensator injects lower currents therefore, reduces losses in the feeder and voltage-source inverter. Further, a saving in the rating of DSTATCOM is achieved which increases its capacity to mitigate voltage sag. Nearly UPF is maintained, while regulating voltage at the load terminal, during load change. A new technique used to regulate the voltage disturbances by state-space model of DSTATCOM with the deadbeat predictive controller. By using this tackle power-quality issues by providing power factor correction, harmonic elimination, load balancing and voltage regulation based on the load requirement. Design and performance evaluation by using MTLAB/SIMULINK.

Keywords: *Distribution Static compensator;, Active Power Filters, Battery Energy Storage Systems, Distribution Series Capacitors, Distribution static compensator, Surge Arresters, Static Electronic Tap Changers, Solid State Transfer Switches, Solid State Fault Current Limiter, Static Var Compensator, Thyristor Switched Capacitors, Uninterruptible Power Supplies.*

[1] Assistant Professor, Department of EEE, Annamacharya Institute of Tech & Sciences, Hyderabad
[2] PG Student, Department of EEE, Annamacharya Institute of Tech & Sciences, Hyderabad
[3] Associate Professor, Department of EEE, Annamacharya Institute of Tech & Sciences, Hyderabad
*Responding Author

Renewable Energy Sources are those energy sources which are not destroyed when their energy is harnessed. Human use of renewable energy requires technologies that harness natural phenomena, such as sunlight, wind, waves, water flow, and biological processes such as anaerobic digestion, biological hydrogen production and geothermal heat. Amongst the above mentioned sources of energy there has been a lot of development in the technology for harnessing energy from the wind. Wind is the motion of air masses produced by the irregular heating of the earth's surface by sun. These differences consequently create forces that push air masses around for balancing the global temperature or, on a much smaller scale, the temperature between land and sea or between mountains. One of the most common power quality problems today is voltage dips. A voltage dip is a short time (10 ms to 1 minute) event during which a reduction in r.m.s voltage magnitude occurs. It is often set only by two parameters, depth/magnitude and duration. The voltage dip magnitude is ranged from 10% to 90% of nominal voltage (which corresponds to 90% to 10% remaining voltage) and with a duration from half a cycle to 1 min. In a three-phase system a voltage dip is by nature a three-phase phenomenon, which affects both the phase-to-ground and phase to-phase voltages. A voltage dip is caused by a fault in the utility system, a fault within the customer's facility or a large increase of the load current, like starting a motor or transformer energizing. Typical faults are single-phase or multiple-phase short circuits, which leads to high currents. The high current results in a voltage drop over the network impedance. At the fault location the voltage in the faulted phases drops close to zero, whereas in the non faulted phases it remains more or less unchanged. Voltage dips are one of the most occurring power quality problems. Off course, for an industry an outage is worse, than a voltage dip, but voltage dips occur more often and cause severe problems and economical losses.

Utilities often focus on disturbances from end-user equipment as the main power quality problems. This is correct for many disturbances, flicker, harmonics, etc., but voltage dips mainly have their origin in the higher voltage levels. Faults due to lightning, is one of the most common causes to voltage dips on overhead lines. If the economical losses due to voltage dips are significant, mitigation actions can be profitable for the customer and even in some cases for the utility. Since there is no standard solution which will work for every site, each mitigation action must be carefully planned and evaluated. There are different ways to mitigate voltage dips, swell and interruptions in transmission and distribution systems. At present, a wide range of very flexible controllers, which capitalize on newly available power electronics components, are emerging for custom power applications. Among these, the distribution static compensator and the dynamic voltage restorer are most effective devices, both of them based on the VSC principle. STATCOM is often used in transmission system. When it is used in distribution system, it is called DSTATCOM STATCOM in Distribution system).

DSTATCOM is a key FACTS controller and it utilizes power electronics to solve many power quality problems commonly faced by distribution systems. Potential applications of D-STATCOM include power factor correction, voltage regulation, load balancing and harmonic reduction. Comparing with the SVC, the DSTATCOM has quicker response time and compact structure. It is expected that the D-STATCOM will replace the roles of SVC in nearly future. D-STATCOM and STATCOM are different in both structure and function, while the choice of control strategy is related to the main-circuit structure and main function of compensators, so D-STATCOM and STATCOM adopt different control strategy. At present, the use of STATCOM is wide and its strategy is mature, while the introduction of D-STATCOM is seldom reported. Many control techniques are reported such as instantaneous reactive power theory (Akagi et al., 1984), power balance theory, etc. In this paper, an indirect current control technique (Singh et al., 2000a, b) is employed to obtain gating signals for the Insulated Gate Bipolar Transistor (IGBT) devices used in current controlled voltage source inverter (CC-VSI) working as a DSTATCOM. This paper considers the operation of DSTATCOM in VCM and proposes a control algorithm to obtain the reference load terminal voltage by using Photovoltaic system. This algorithm provides the combined advantages of CCM and VCM. The UPF operation at the PCC is achieved Rsh - shunt resistance; I-V characteristic of a PV module is highly non-linear in nature. This characteristics drastically changes with respect to changes in the solar radiation and cell temperature.

Whereas the solar radiation mainly affects the output current, the temperature affects the terminal voltage. Fig.2 shows the I-V characteristic of the PV module under varying solar radiations at constant cell temperature (T = 25 °C). Fig 2: Current versus voltage at constant cell temperature T = 25 °C. Fig.3 shows the I-V characteristics of the PV module under varying cell temperature at constant solar radiation (1000 W/m2). Fig 3: Current versus voltage at constant solar radiation G = 1000 W/m2 III. PROPOSED CONTROL SCHEME: Circuit diagram of a DSTATCOM -compensated distribution system is shown in Fig. 4. It uses a three- phase, four wire, two-level, neutral-point-clamped VSI. This structure allows independent control to each leg of the VSI. Fig. 5 shows the single-phase equivalent representation of Fig. 1. Variable is a switching function, and can be either or depending upon switching state. Filter inductance and resistance are and, respectively. Shunt capacitor eliminates high-switching frequency components. First, discrete modeling of the system is presented to obtain a discrete voltage control law, and it is shown that the PCC voltage can be regulated to the desired value with properly chosen parameters of the VSI. Then, a procedure to design VSI parameters is presented. A proportional-integral (PI) controller is used to regulate the dc capacitor voltage at a reference value. Based on instantaneous symmetrical component theory and complex Fourier transform, a reference voltage magnitude generation scheme is proposed that provides the advantages of CCM at nominal load. The overall controller block diagram is shown in Fig. 6. These steps are explained as follows. at nominal load, whereas fast voltage regulation is provided during voltage disturbances. Also, the reactive and harmonic component of load current is supplied by the compensator at any time of operation. The deadbeat predictive controller is used to generate switching pulses. The control strategy is tested with a three-phase four-wire distribution system. The effectiveness of the proposed algorithm is validated through detailed simulation and experimental results.

PHOTOVOLTAIC (PV) MODULE

To understand the PV module characteristics it is necessary to study about PV cell at first. A PV cell is the basic structural unit of the PV module that generates current carriers when sunlight falls on it. The power generated by these PV cell is very small. To increase the output power the PV cells are connected in series or parallel to form PV module. The electrical equivalent circuit of the PV cell is shown in Fig 1.

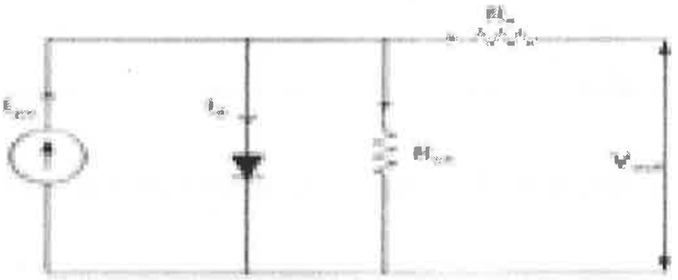


Fig 1: Electrical equivalent circuit diagram of PV cell

The main characteristics equation of the PV module is given by

$$I = I_{pv} - I_o\left[\exp\left(\frac{q(V+IR_s)}{\alpha KT}\right)-1\right] - \frac{V+IR_s}{R_{sh}} \qquad (1)$$

$$I_o = I_{o,n}\left(\frac{T_n}{T}\right)^3 \exp\left[\frac{qE_g}{\alpha K}\right]\left(\frac{1}{T_n}-\frac{1}{T}\right) \qquad (2)$$

$$I_{pv} = [I_{sc} + K_i(T-T_n)]\frac{G}{G_n} \qquad (3)$$

Where,

1. I and V - cell output current and voltage;
2. Io - cell reverse saturation current;
3. T - Cell temperature in Celsius;
4. K - Boltzmann's constant; q -
5. Electronic charge;
6. Ki- short circuit current/temperature coefficient;
7. G - Solar radiation in W/m2;
8. Gn- nominal solar radiation in W/m2;
9. Eg - energy gap of silicon;
10. Io,n - nominal saturation current;
11. Rs - Series resistance;
12. Rsh - shunt resistance;

I-V characteristic of a PV module is highly non-linear in nature. This characteristics drastically changes with respect to changes in the solar radiation and cell temperature.

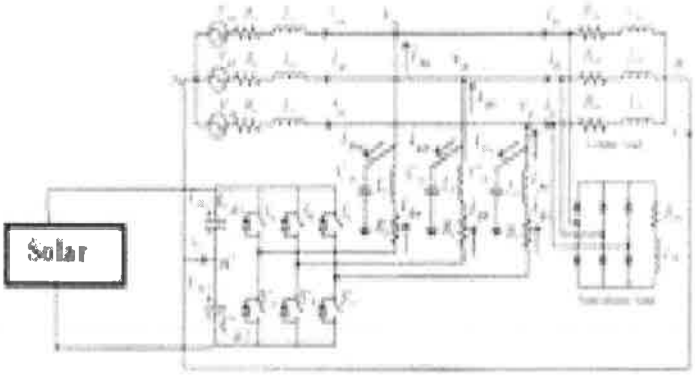


Fig 4: Circuit diagram of the DSTATCOM-compensated distribution system.

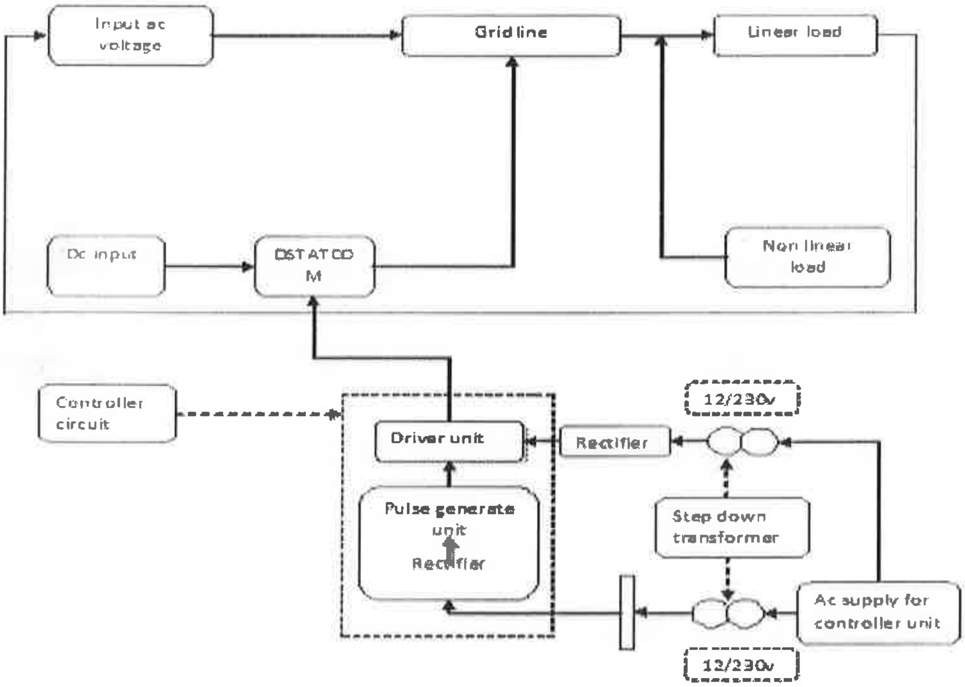


Fig.2 proposed block diagram System Modeling and Generation of the Voltage-Control Law

The state-space equations for the circuit shown in Fig. 5 are given by

$\dot{x} = Ax + Bz$ (4)

where,

$$A = \begin{bmatrix} 0 & \frac{1}{C_{fc}} & 0 \\ \frac{1}{L_f} & \frac{R_f}{L_f} & 0 \\ -\frac{1}{L_s} & 0 & -\frac{R_s}{L_s} \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & -\frac{1}{C_{fc}} & 0 \\ \frac{V_{dc}}{L_f} & 0 & 0 \\ 0 & 0 & \frac{1}{L_s} \end{bmatrix},$$

$$x = [v_{fc} \quad i_{ft} \quad i_s]^t, \quad z = [u \quad i_{Lt} \quad v_{st}]^t$$

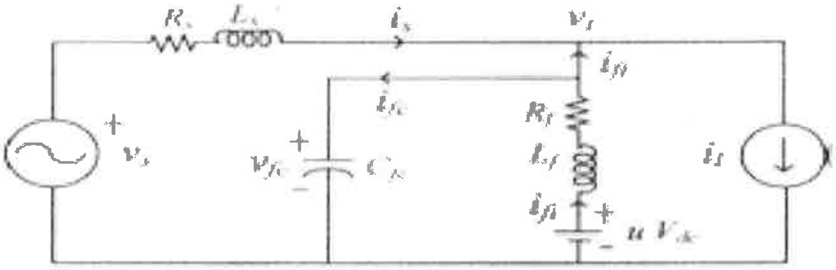


Fig 5: Single-phase equivalent circuit of DSTATCOM

The general time-domain solution of (4) to compute the state vector x(t) with known initial value x(t0) , is given as follows:

$$x(t) = e^{A(t-t_0)}x(t_0) + \int_{t_0}^{t} e^{A(t-\tau)} B\, z(\tau)\, d\tau \qquad (5)$$

The equivalent discrete solution of the continuous state is obtained by replacing and as follows:

$$x(k+1) = e^{AT_d}\, x(k) + \int_{kT_d}^{kT_d+T_d} e^{A(kT_d+kT_d-\tau)} B\, z(\tau)\, d\tau$$

(6)

In (6), k and Td represent the Kth sample and sampling period, respectively. During the consecutive sampling period, the value of Z(t) is held constant, and can be taken as Z(k). After simplification and changing the integration variable, (6) is written as

$$x(k+1) = e^{AT_d}\, x(k) + \int_{0}^{T_d} e^{A\lambda} B\, d\lambda\; z(k). \qquad (7)$$

Equation (6) is rewritten as follows:

$$x(k+1) = G\, x(k) + H\, z(k) \qquad (8)$$

Where, G and H are sampled matrices, with a sampling time of Td. For small sampling time, matrices G and H are calculated as follows:

$$G = \begin{bmatrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} & G_{23} \\ G_{31} & G_{32} & G_{33} \end{bmatrix} = e^{AT_d} \approx I + AT_d + \frac{A^2T_d^2}{2}$$
$$H = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} = \int_{0}^{T_d} e^{A\lambda} B\, d\lambda$$
$$\approx \int_{0}^{T_d} (I + A\lambda)\, B\, d\lambda. \qquad (9)$$

From (9),
$G_{11} = 1 - T_d^2/2L_fC_{fc}$, $G_{13} = 0$, $H_{11} = T_d^2V_{dc}/2L_fC_{fc}$, Hence, the capacitor voltage using (8) is given as

$$v_{fc}(k+1) = G_{11}v_{fc}(k) + G_{12}i_{fi}(k) + H_{11}u(k) + H_{12}i_{ft}(k)$$

(10)

As seen from (8), the terminal voltage can be maintained at a reference value depending upon the VSI parameters Vdc, Cfc, Lf , Rf , and sampling time Td . Therefore, VSI parameters must be chosen carefully. Let Vt* be the reference load terminal voltage. A cost function is chosen as follows

$$J = [v_{fa}(k+1) - v_t^*(k+1)]^2 \quad (11)$$

The cost function is differentiated with respect to u(k) and its minimum is obtained at

$$v_{fa}(k+1) = v_t^*(k+1) \quad (12)$$

The deadbeat voltage-control law, from (10) and (12), is given as

$$u^*(k) = \frac{v_t^*(k+1) - G_{11}v_{fa}(k) - G_{12}i_{fa}(k) - H_{12}i_{la}(k)}{H_{11}}$$

(13)

In (13), u*(K−1) is the future reference voltage which is unknown. One-step-ahead prediction of this voltage is done using a second-order Lagrange extrapolation formula as follows:

$$v_t^*(k+1) = 3v_t^*(k) - 3v_t^*(k-1) + v_t^*(k-2). \quad (14)$$

The term u*(K+1) is valid for a wide frequency range and when substituted in (13), yields to a one-step-ahead deadbeat voltage-control law. Finally u*(k) is converted into the ON/OFF switching command to the corresponding VSI switches using a deadbeat hysteresis controller

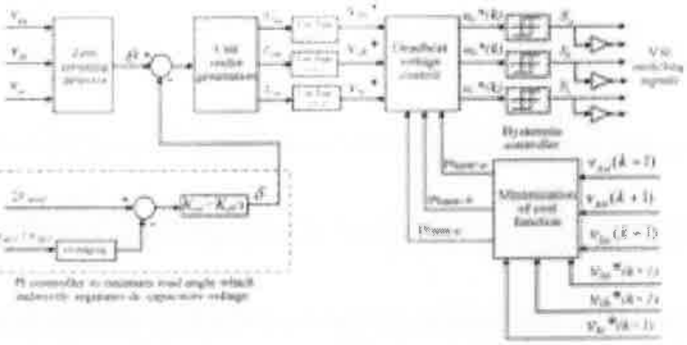


Fig 6: Overall block diagram of the controller to control DSTATCOM in a distribution system.

SIMULATION RESULTS

The below figure shows that simulation circuit diagram and the respective waveforms as shown with the MATLAB/Simulink software.

Fig 7: Simulation circuit diagram of a proposed converter system

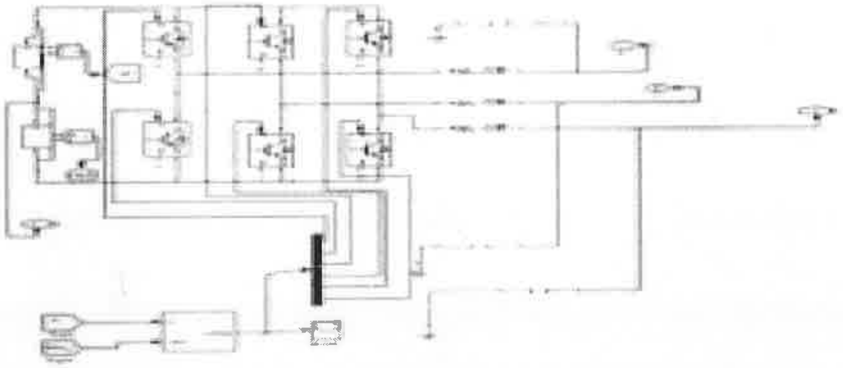


Fig 8: Simulation circuit diagram of a shunt active power filter

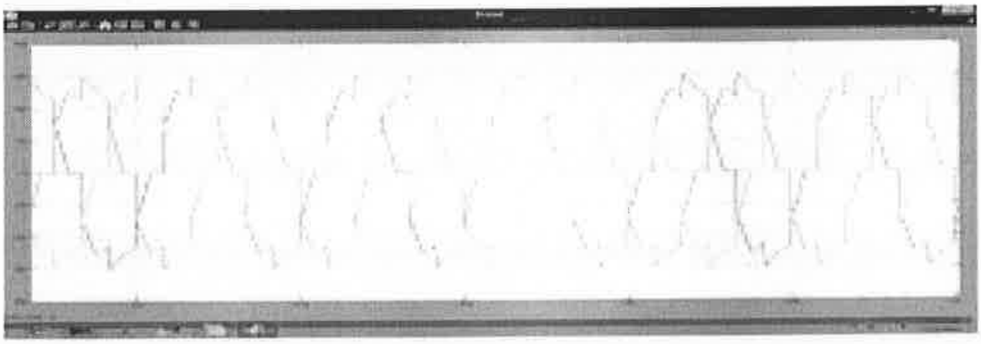


Fig 9 Before compensation (a)Terminal voltages

Under normal operation conditions with out Dstatcom terminal voltages unbalanced conditions

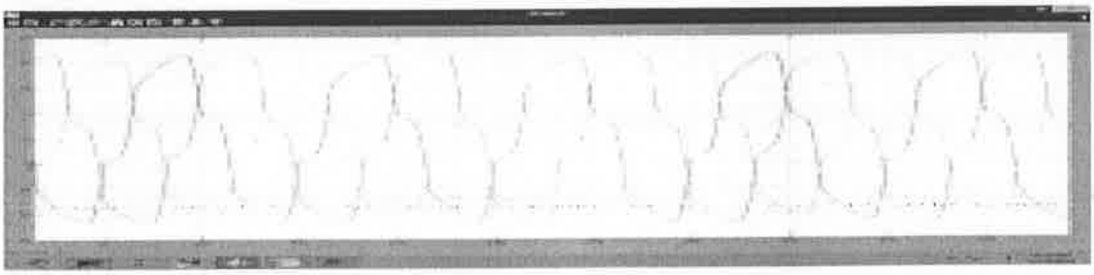


Fig 10 Before compensation (b) Source currents

Under normal operation condition with out Dstatcom source current unbalanced condition,

Fig 11 Terminal voltages and source currents using the proposed method phase(a)

The impact of load changes on system performance; load is increased to 140%of its nominal value. Under this condition, the traditional method gives less power factor as the compensator will supply more reactive current to maintain the reference voltage. The voltage and current waveforms.

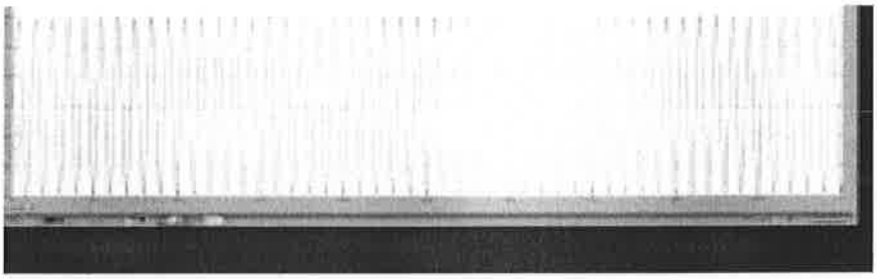


Fig 12 Terminal voltages and source currents using load change in proposed method

The proposed method is considered. Fig. 6.11 shows the regulated terminal voltages and corresponding source currents in phases and respectively.

Fig 13 Terminal voltages and source currents using the traditional method (c) Phase

CONCLUSION

The performance of the proposed scheme is compared with the traditional voltage controlled DSTATCOM. The proposed method provides the following advantages- at nominal load, the compensator injects reactive and harmonic components of load currents, resulting in UPF; nearly UPF is maintained for a load change; fast voltage regulation has been achieved during voltage disturbances and losses in the VSI and feeder are reduced considerably, and have higher sag supporting capability with the same VSI rating compared to the traditional scheme. Different types of voltage sag conditions should applied compensated in simulink environment. Additionally power factor correction and voltage regulation the harmonics are also checked, 20% voltage sag eliminated under t=0.5 to 1sec, thus the simulation results show that the proposed scheme provides DSTATCOM, a capability to improve several Power Quality problems (related to voltage and current).

Acknowledgments
The author appreciates all those who participated in the study and helped to facilitate the research process.

Conflict of Interests: The author declared no conflict of interests.

REFERENCES

M. Bollen, Understanding Power Quality Problems. Piscataway, NJ, USA: IEEE, 2000, ch. 1, pp. 1–35.

H. Fujita and H. Akagi, "Voltage-regulation performance of a shunt active filter intended for installation on a power distribution system," IEEE Trans. Power Electron., vol. 22, no. 3, pp. 1046–1053, May 2007.

A. Ghosh and G. Ledwich, "Load compensating DSTATCOM in weak ac systems," IEEE Trans. Power Del., vol. 18, no. 4, pp. 1302–1309, Oct. 2003.

A. Elnady and M. Salama, "Unified approach for mitigating voltage sag and voltage flicker using the DSTATCOM," IEEE Trans. Power Del., vol. 20, no. 2, pt. 1, pp. 992–1000, Apr. 2005.

S. Rahmani, A. Hamadi, and K. Al-Haddad, "A Lyapunov-functionbased control for a three-phase shunt hybrid active filter," IEEE Trans. Ind. Electron., vol. 59, no. 3, pp. 1418–1429, Mar. 2012.

M. K. Mishra and K. Karthikeyan, "A fast-acting dclink voltage controller for three-phase DSTATCOM to compensate ac and dc loads," IEEE Trans. Power Del., vol. 24, no. 4, pp. 2291–2299, Oct. 2009.

M. K. Mishra, A. Ghosh, A. Joshi, and H. M. Suryawanshi, "A novel method of load compensation under unbalanced and distorted voltages,"IEEE Trans. Power Del., vol. 22, no. 1, pp. 288–295, Jan. 2007.

M. K.Mishra, A. Ghosh, and A. Joshi, "Operation of a DSTATCOM in voltage control mode," IEEE Trans. Power Del., vol. 18, no. 1, pp. 258–264, Jan. 2003.

A. Jain, K. Joshi, A. Behal, and N. Mohan, "Voltage regulation with STATCOMs:Modeling, control and results," IEEE Trans. Power Del.,vol. 21, no. 2, pp. 726– 735, Apr. 2006.

R. Gupta, A. Ghosh, and A. Joshi, "Switching characterization of cascaded multilevel-inverter-controlled systems," IEEE Trans. Ind. Electron., vol. 55, no. 3, pp. 1047–1058, Mar. 2008.

Author's Details

M.Chandra Shekhar completed his B.Tech in Electrical and electronics Engineering from PRRM Engineering college(JNTUH),shabad in the year 2005 and received M.Tech in the stream of power electronics Engineering at AURORAS engineering college (JNTUH),Bhongir, Nalgonda Dist,in the year 2012.pursuing PhD in veltech Dr.RR & DR.SR university Chennai. Currently working as an Assistant Professor in Annamacharya Inst.of Tech & sciences, Hyderabad, since2012, and his areas of interests are Micro Grid, Renewable energy sources.

P. Renuka Presently pursuing M.Tech in the Dept of EEE (PE), Annamacharya Institute of Technology And Sciences, Hyd, Ts, India.
E mail Id: renuka8892@gmail.com

Mr. Jadapalli Sreedhar was born at rajampet near Kadapa(Dist) AP., India. He completed his B.Tech in Electrical and Electronics Engineering from JNTU in the year 2002. He completed his M.Tech in power electronics from JNTU in the year 2006. Currently, he is pursing PhD from the GITAM University, Hyderabad campus on the topic synchronous buck convertor applications. He is working as associate professor in the Dept of EEE, Annamacharya Institute of Technology and Sciences, Hyderabad since 2012. His fields of interest are power systems and power electronics. So far he has published 15 papers in international Journals and 4 papers in national Journal.
Email Id: sreedharmtech@gmail.com

International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 4 Issue 3, March-2015, pp: (67-75), Impact Factor: 1.252, Available online at: www.erpublications.com

Power Quality Improvement for Microgrid with Multiple Energy Sources

Mr. Kuldip Singh[1], Mr. M. Chandrashekhar[2], Mrs. P. Swathi[3]
GNIT[1], AITS[2]

Abstract: The hybrid micro gird is growing due to the potential benefits in providing safe, reliable and sustainable electricity from renewable energy sources. In this paper, we are discussing the multiple energy sources i.e. PV/Fuel cell and wind based hybrid microgrid with help of simulation approach to improve the power quality in microgrid with ANN control. The artificial neural network control is used for dc/ac conversion with PWM and DVR control to improve the power quality. A small hybrid microgrid has been modeled and simulated using the simulink in the MATLAB. The simulation results are comparing with Pi control.

Keywords: microgrid, hybrid, renewable energy sources, PV, Fuel cells, wind, Power quality, ANN control, Simulation, Pi control.

I. INTRODUCTION

In recent years the microgrid has been growing due to its several potential and economic advantages like, the mircogrid has small investment, which reduces capital exposure and risk by closely matching capacity increases to growth in demand. It is also reduced transmission and distribution cost. The micro-grid has less energy losses and higher overall energy efficiency [1].

The progressive decrease of fossil fuels like coal, diesel and increase the environmental problem associated to their combustion force to search the alternative sources of electrical energy to meet the load demand [2]. The micro-grid is designed based on the renewable energy sources near by the Load. The fuel cells, PV cell technology and wind power is mainly used in the microgrid to full-fill the load demand with hybrid technology. In hybrid micro-grid the power quality problem is the measures aspect, which is affecting the load demand. The main cause of power quality problem is power electronic components, which required in microgrid to convert dc to ac. The output of the inverter in microgrid should be compatible in voltage and frequency with load [3].

In ac micro-grid dc power from photovoltaic panels and fuel cells has been converted into ac using dc/dc boosters and dc/ac inverter in order to connect the ac load. Due to static devices in converter results are harmonics injection and lower power factor to electric power system [6]. The load equipments of the modern generation are more sensitive. Due to harmonics can initiate production loss, economic loss and environmental effect [10]. In this paper, we are discussing the dc/ac inverter based on the ANN control to improve the performance of PWM inverter. To minimize the harmonics in the microgrid the ANN based DVR control is proposed.

II. CONTROLS IN MICROGRID

In the microgrid the major source of electricity are Photovoltaic cells and Fuel cells. These sources are generating the dc output. The load connected to the grid is ac load. The conversion devices have main role in the microgrid.

a) **Conversion control:** The Photovoltaic cells and fuel cells are connected to the dc bus through a boost DC/DC converter in order to generate the maximum power from PV cells and fuel cells.. After boost the dc the load is connected to the grid. The PWM based inverter is used to convert the dc to ac. To operate or control the boost converter and inverter different gate control are used for accurate output and to minimize the power losses . The major problem with control devices are harmonic generation[3].

b) **Power quality control:** The power electronics devices are used for interface the renewable sources to microgrid or interconnected with other sources like renewable and non-renewable generators, storage systems and load in microgrid. The microgrid is different from the main grid, where the large and sudden changes in the load may results in voltage transient of large magnitudes in ac bus. The non-linear loads and switching power converters are decreasing the power quality in microgrid [5]. To overcome the power quality problem in distribution the DVR control is proposed.

The apparent Power

$S_{DVR} = I_{Load}V_{DVR} = I_{Load}(V_{Load} - V_s)$ (5)

Active Power

$P_{DVR} = I_{Load}V_{DVR}cos\theta_s = I_{Load}(V_{Load} - V_s)cos\theta_s$ (6)

The magnitude and angle of DVR voltage

$V_{DVR} = V_{Load} - V_S$ (7)

$\theta_{DVR} = \theta_s$ (8)

c) **In-Phase Advanced compensation (IPAC):** This method is controlling injection energy, in phase advance compensation method was proposed. The injection of active power is made zero by means the injection voltage phasor perpendicular to the load current phasor. This method reduce the energy stored in DC link by injecting reactive power [14].

IV. ANN CONTROL FOR MICROGRID

To improve the performance of the gate control circuit for PWM inverter and DVR control in PV/Fuel cell based microgrid, a multilayer back propagation type artificial neural network controller is used. The back propagation algorithm is used to train the network. The Gradient decent method is used to find the local minimum of a given function. The GD method is the first order optimization algorithm and it is robust when it start far of the final minimum. The Levenberg Marquardt back propagation algorithm is the second order optimization and it is more robust & finds a solution even if it does begin far from the final optimum. The Levenberg Marquardt algorithm is interpolates between the Gauss Newton algorithm and gradient decent method and it is best comparing to Gauss Newton and gradient decent method.[10].

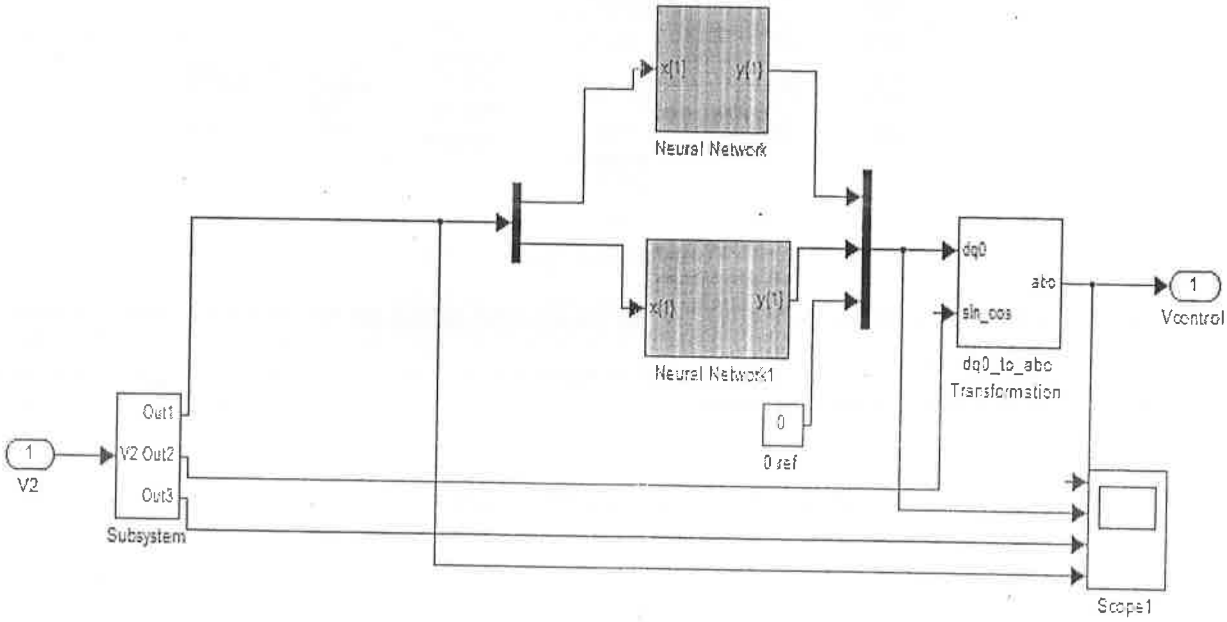


Fig:-3: ANN control circuit for Inverter control

In fig-3 shown the artificial neural network control simulation circuit in MATLAB simulation for DVR control and PWM inverter control in the microgrid with Levenberg Marquardt back propagation algorithm. All the input are used to train the ANN from conventional controller. As shown Fig-4 the artificial neural network contain the three layer composed of two input layers and one output layer. Here Input1 and a(1) are the input layers and a(2) are the output layer of the network. Each input layer have the input with weights W_{11} with adder function to compute the weighted sum and input of the layer. It is also containing a Linear transfer function as activation function and bias b.

Output= activation function (weighted sum of inputs + bias) (9)

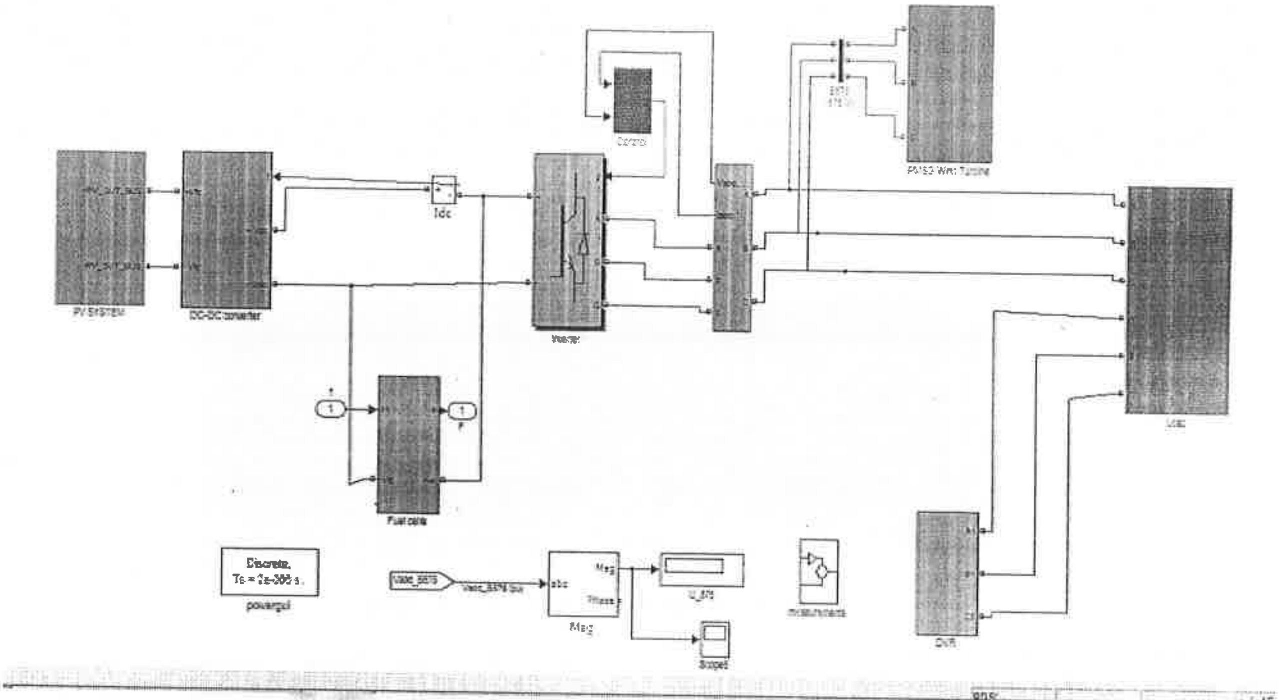


Fig: 6: ANN control for microgid with multiple sources

As shown in Fig-7 Input voltage for control circuit feeding from the load end to control circuit. The input signal will contain the high value of harmonic as shown in Fig-8. As per the FFT analysis the harmonics value is 41.49%.

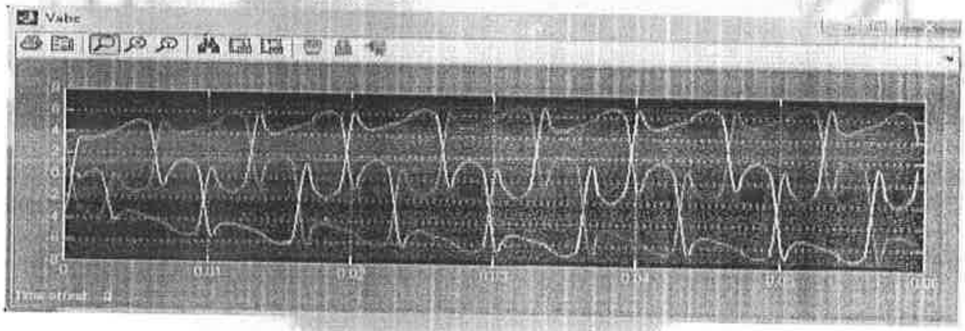


Fig. 7: Input voltage to control circuit

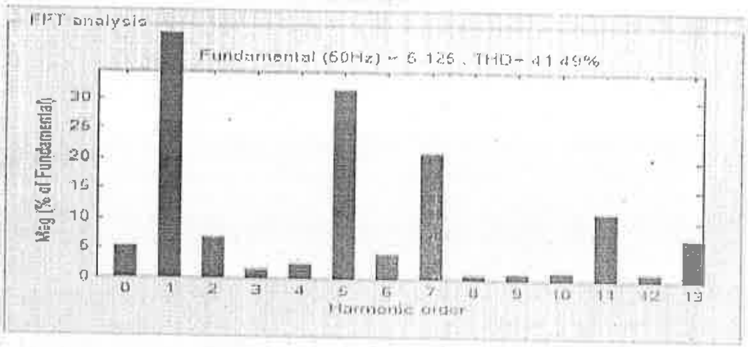


Fig. 8: FFT analysis for Input signal of the control circuit

The control signal will feed to the PWM generator to generate the gate pulses for controlling PV inverter in microgrid.

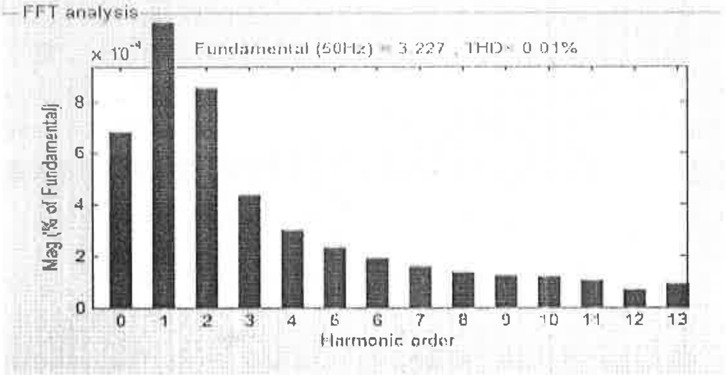


Fig:-12: FFT analysis for Vcontrol signal

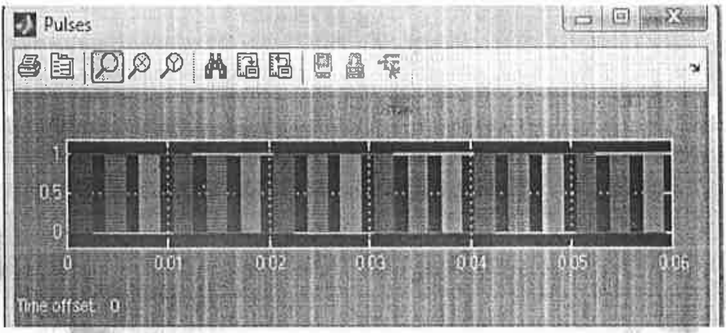


Fig:-13: Gate pulses for gate control

The voltage and current wave form are shown in the Fig-14 and Fig:-15 for PV & Fuel Cell based microgrid. The FFT analysis for voltage and current with ANN DVR control shown in Fig:-16 and Fig-17 and voltage & current waveform with Pi control are shown in Fig:18 and Fig 20.The FFT analysis for current and voltage with Pi control shown in Fig 19 and Fig-21.

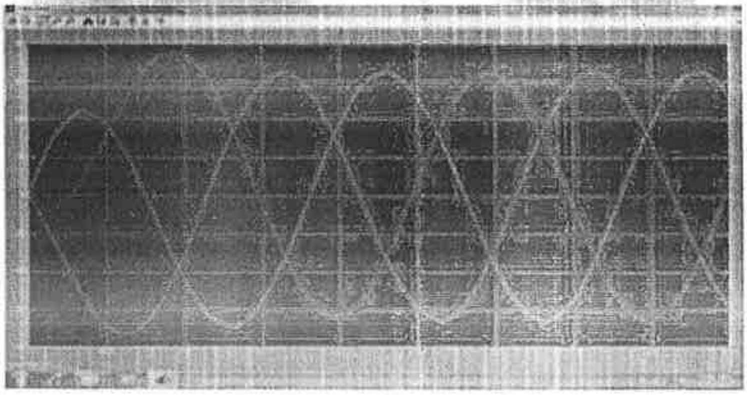


Fig:-14: Load voltage with ANN based DVR control

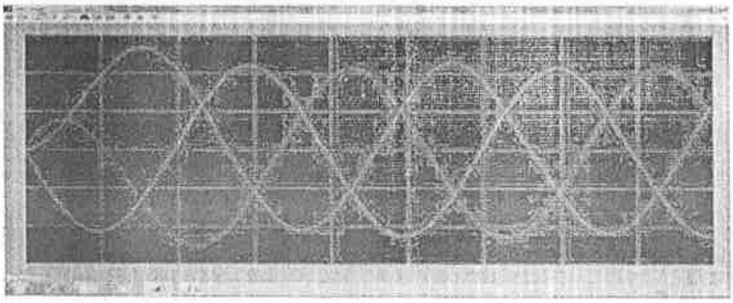


Fig:-15: Load current with ANN based DVR control

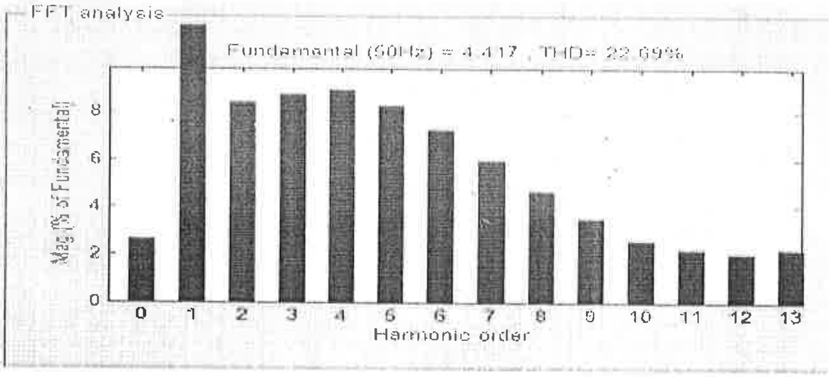


Fig: 21: Load current with Pi control based DVR control

VI. Conclusion

The paper present PV inverter & DVR control with artificial neural network and Pi control to convert the dc to ac in PV & Fuel Cell based microgrid. The artificial neural control used the reference voltage to generate PWM switching signals. The ANN control will remove the error from reference signal and improve the switching signals. The ANN control, training algorithm and principal operation for inverter control were analyzed in details. The DVR control are improving the power quality in microgrid. Experimental results indicate that the THD with ANN control signals and Pi control signals for PWM inverter and DVR. From the result analysis ANN control signals THD value is much less than that of conventional method.

References

[1]. Mohammad mohammadi, "Review of simulation and optimization of Autonomous and grid connected hybrid renewable energy systems as micro-grids", ISESCO,Vol-9,PP-60-67.

[2]. J.I.San Martin, I.Zamora,J.J san Martin,V.Aperribay. P.Eguia "Hybrid Technologies: Fuel Cells and Renewable Energies" University of Basque Country.

[3]. Ahmad Eid, " Performance of Grid-connected Hybrid photovoltaic/ Fuel cell/Battery Distributed Generation Systems", International conference on Electrical Engineering and Computer Sciences,March 15-17 2013, Japan,pp-147-154.

[4]. Xiong Liu,Peng Wang and Poh Chiang Loh, " A hybrid AC/DC Microgrid and Its coordination control", IEEE Transaction on smart Grid,Vol-2,No-2,June-2011.

[5]. Gelu Gurguiatu,IonelVechiu,Toader Munteanu, " Power quality improvement using renewable energy",

[6]. Georgios A. Tsengenes and Georgios A.Adamidis, " Performance Evaluation of Pi and Fuzzy controlled power Electronic Inverters for Power quality improvement",Chapter-22 INTECH.

[7]. Ngac Ky Nguyen,Patrice Wira, Damien Flieller,Djaffar Ould Abdeslam,Jean Merckle, " A comparative experimental study of neural and conventional controllers for an Active power filer", IECON10.

[8]. Fahad Ali,Dr.Abdul Aziz Bhatti,Mashood Nasir, M.Arif Saeed, "FuelCell based intelligent hybrid energy storage system and grid integration", Proceeding of International conference on Energy and Sustainabillity-2013.

[9]. Frede Blaabjerg, Remus Teodorescu, Zhe Chen,MarcoLiserre, " Power Converter and control of Renewable Energy Systems", Aalborg University.

[10]. C.K.Sundarabalan and K.Selvi, "Power Quality Enhancement in Power Distribution system using artificial intelligence based Dynamic voltage Restorer", IJEEI-Vol-2,No-4,Dec-2013.

[11]. R.H,Salimin, "Simulation analysis of DVR Performance for Voltage sag mitigation", IEEE power engg. And optimization conference,Malaysia,pp 261-266,june 2011.

[12]. Fuel cell system Explained,James Larminie and Andrew Dicks,2nd Edition,Wiley,ISBN 0-470-84857-X.

[13]. John Newman,M. D.Grahame Holmes, J. Godsk Nielsen and F. Blaabjerg, " A dynamic voltage restorer (DVR) with selective harmonics compensation at medium voltage level", IEEE-2013.

[14]. C.Benachaiba and B.Ferdi, " Power Quality improvement using DVR",American Journal of Applied Sciences 6(3): 396-400 ISSN 1546-9239.

[15]. 1 M.Narendra Kumar,.K.S.R. Anjaneyulu , 3Kuldip singh." Power Quality Improvement in Psmg Based Microgrid with 12-Pulse Converter" IEEJ Vol-4(2013) PP 1079-1086 ,ISSN 2078-2365.

Power Quality Improvement by Voltage Control Using Dstatcom

1.PODHILA RENUKA,

1.PG SCHOLAR ,DEPT OF EEE(EPS) ,ANNAMACHARYA INSTITUTEOF TECHNOLOGY AND SCIENCES, HYD, TS,INDIA

2.CHANDRA SHEKHAR

2. ASSISTANT PROFESSOR, DEPT OF EEE ANNAMACHARYA INSTITUTE OF TECHNOLOGY AND SCIENCES, HYD,TS, INDIA

ABSTRACT This paper proposes a new topology by Distribution Static Compensator using Matlab. This proposed method of power quality improvement achieves UPF which is not possible in previous methods. Maximum UPF is maintained, while regulating voltage at the load terminal, during fluctuation of load. Dstatcom solves Power quality issues by achieving PF correction, harmonic elimination, load balancing, and voltage regulation based on the load requirement. Keywords: power quality, DSTATCOM, PF correction, harmonic elimination, load balancing, voltage control, matlab.

INTRODUCTION In recent years, Electrical Power Quality had obtained more attention in power engineering. In present day's power distribution system is suffering from severe power quality problems. These power quality problems include high reactive power burden, harmonics currents, load unbalance, excessive neutral current etc. The measure of power quality depends upon the needs of the equipment that is being supplied [1]. What is good power quality for an electric motor may not be good enough for a personal computer. Usually the term power quality refers to maintaining a sinusoidal waveform of bus voltages at rated voltage and frequency. The waveform of electric power at generation stage is purely sinusoidal and free from any distortion. Many of the power conversion and consumption equipment are also designed to function under pure sinusoidal voltage waveforms. However, there are many devices that distort the waveform. These distortions may propagate all over the electrical network. In recent years, there has been an increased use of non-linear loads which has resulted in an increased fraction of non-sinusoidal currents and voltages in Electric Network. A Distribution System Suffers from Current as well as voltage related Power Quality Problems, which include poor power factor, distorted source current and voltage disturbances [2]. DSTATCOM are used in the distribution system for improvement of power quality issues. The voltage sags/swells have become the main cause of equipment malfunctioning, tripping in the industries due unbalance between the power supply and demand. From the last decade, there have been considerable developments and improvements in energy storage technologies [12]. This paper considers the operation of DSTATCOM in VCM and proposes a control algorithm to obtain reference load terminal voltage. This algorithm provides both advantages of VCM and CCM.UPF operation is achieved at nominal load, whereas fast voltage regulation is provided during voltage fluctuations. At the same time reactive and harmonic component of load current is supplied by the compensator at any time of Operation. The entire control is tested with three phase four wire distribution system. This proposed algorithm is validated through simulation and experimental results.

POWER INJECTION PRINCIPLE The total apparent (complex) power that is injected into a transmission line is made up of two components, namely active and reactive. The active power P component is the part of energy that is converted into physical energy form. The reactive power Q component helps create the indispensable magnetic medium needed for most of today's electromagnetic energy conversion devices and systems. The majority of industrial and commercial appliances require both active and reactive power components for operation. Both P and Q are needed instantly and in different quantities to meet the requirement of the electrical energy converting device connected to the AC source [3]. .Reactive power can be absorbed or supplied depending on the energy medium associated with the electric device. Energy absorbing or supplying components are reactors and capacitors respectively. Reactors absorb reactive power +Q and draw lagging current [15]. The consumed energy is stored as a magnetic energy in the reactor turns. Meanwhile, capacitors supply reactive power -Q and draw leading current, storing it as electric charge within its dielectric medium and associated charge plates. To understand P and Q flow in a transmission system, consider a simple system that is made up of sending and receiving buses with a transmission cable in between as shown in Figure 1 [13, 14].

BLOCK DIAGRAM OF PROPOSED SYSTEM Figure-2 represents the block diagram of the proposed system. DSTATCOM regulates terminal voltage satisfactorily; depending upon the properly chosen VSI parameters. AC source is excited with three phase voltage. Controller is activated with 5V and the drive amplifier is activated with 12V DC supply. This amplifier enhances the input values to the DSTATCOM.

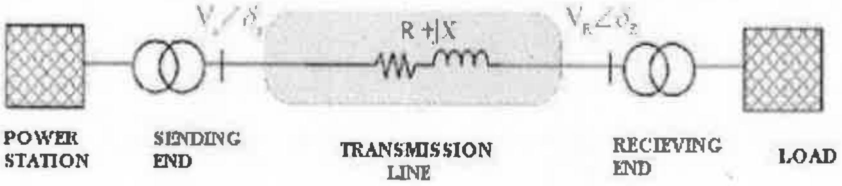


Figure-1. Transmission system.

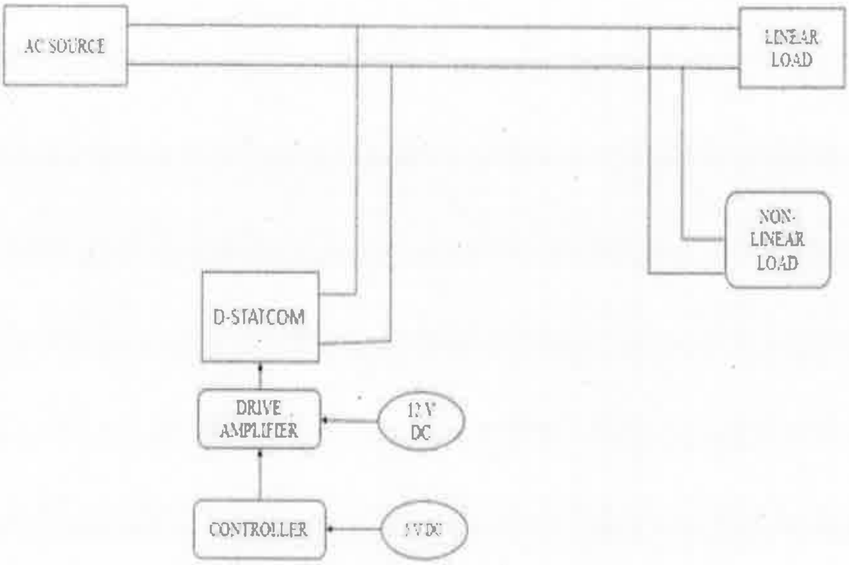


Figure-2. Block diagram.

VSI PARAMETERS DESIGN The Dc bus voltage is taken twice the peak of phase voltage of source value. Value of DC capacitors are chosen based on a period of Sag/Swell and change in DC bus voltage during transients [4]. This voltage value continues to decrease until the capacitor voltage controller comes into action. Inductance Filter provides reasonably high switching frequency and sufficient rate of change of current so that VSI currents follow desired currents. PROPOSED METHOD USING LINEAR LOAD This control scheme is implemented using Matlab

2014 software. Distorted and unbalanced source currents flowing through the feeder make terminal voltages unbalanced and distorted. Simulation parameters are mentioned below as,

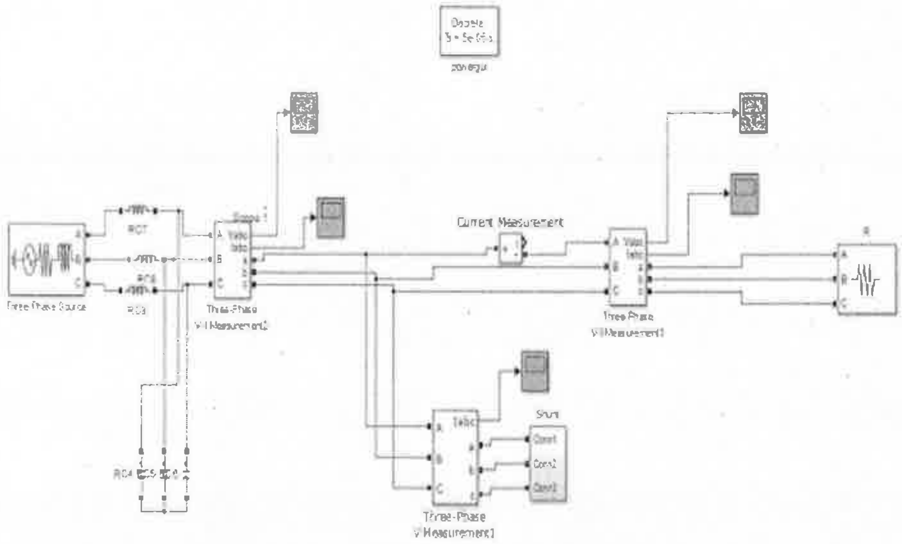


Figure-3. Simulation diagram for linear load with DSTATCOM.

The proposed method is experimentally verified on a reduced scale set up. In this method the rms value of source current is reduced from 0.61 to 0.73 A and source current is also reduced to 1.75 A from 1.84 A in this proposed method [7, 8]. Thereby losses in VSI are reduced and also capability of DSTATCOM to mitigate deep sag is increased. Hence the proposed scheme is able to provide fast voltage regulation. The experimental results are quite consistent with the simulation results. They prove the effectiveness of the proposed control system.

CONCLUSIONS In this paper, a method has been proposed for the generation of reference load voltage for a voltagecontrolled DSTATCOM [9]. The performance of the proposed scheme is compared with the traditional voltagecontrolled DSTATCOM using linear and nonlinear load [11]. The proposed method satisfies the following conditions such that maintenance of UPF even at load changes, better voltage regulation, losses are reduced in VSI. The simulation and experimental results how that the proposed scheme provides DSTATCOM, a capability to improve several PQ problems.

REFERENCES

[1] M. Bollen. 2000. Understanding Power Quality Problems. Piscataway, NJ, USA: IEEE. 1: 1-35.

[2] H. Fujita and H. Akagi. 2007. Voltage-regulation performance of a shunt active filter intended for installation on a power distribution system. IEEE Trans. Power Electron. 22(3): 1046-1053.

[3] A. Ghosh and G. Ledwich. 2003. Load compensating DSTATCOM in weak ac systems. IEEE Trans. Power Del. 18(4): 1302-1309.

[4] A. Elnady and M. Salama. 2005. Unified approach for mitigating voltage sag and voltage flicker using the DSTATCOM. IEEE Trans. Power Del. 20(2), pt. 1: 992-1000.

[5] S. Rahmani, A. Hamadi and K. Al-Haddad. 2012. A Lyapunov-function based control for a three-phase shunt hybrid active filter. IEEE Trans. Ind. Electron. 59(3): 1418-1429.

[6] M. K. Mishra and K. Karthikeyan. 2009. A fast-acting dc-link voltage controller for three-phase DSTATCOM to compensate ac and dc loads. IEEE Trans. Power Del. 24(4): 2291-2299. [7] M. K. Mishra, A. Ghosh, A. Joshi, and H. M. Suryawanshi. 2007. A novel method of load compensation under unbalanced and distorted voltages. IEEE Trans. Power Del. 22(1): 288-295. [8] M. K. Mishra, A. Ghosh and A. Joshi.2003. Operation of a DSTATCOM in voltage control mode. IEEE Trans. Power Del. 18(1): 258-264. A. Jain, K. Joshi, A. Behal, and N. Mohan. 2006. Voltage regulation with STATCOMs: Modeling, control and results. IEEE Trans. Power Del. 21(2): 726-735.

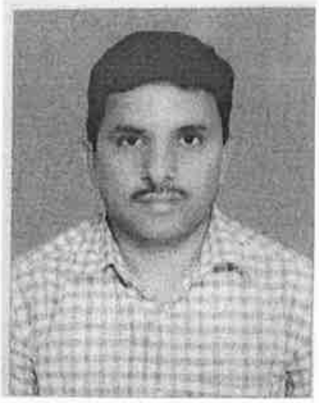
[9] R. Gupta, A. Ghosh and A. Joshi. 2008. Switching characterization of cascaded multilevel-invertercontrolled systems. IEEE Trans. Ind. Electron. 55(3): 1047-1058.

[10]P. Mitra and G. Venayagamoorthy. 2010. An adaptive control strategy for DSTATCOM applications in an electric ship power system. IEEE Trans. Power Electron. 25(1): 95-104.

AUTHOR'S DETAILS:

PODHILA RENUKA,

PG SCHOLAR ,DEPT OF EEE(EPS) ,ANNAMACHARYA INSTITUTEOF TECHNOLOGY AND SCIENCES, HYD, TS,INDIA



CHANDRA SHEKHAR
ASSISTANT PROFESSOR, DEPT OF
EEE ANNAMACHARYA INSTITUTE OF
TECHNOLOGY AND SCIENCES,
HYD,TS, INDIA

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

Performance Improvement of Renewable Energy Systems through Active Power Filters for Meeting the Energy Demand and Power Quality Improvement

M.B.Hemanth Kumar[1],M.Chandrashekhar[2]

Assistant Professor,Dept. Of EEE, Annamacharya Institute Of Technology And Sciences,Hyderabad,India[1]

Assistant Professor,Dept. Of EEE, Annamacharya Institute Of Technology And Sciences,Hyderabad,India[2]

ABSTRACT: Renewable generation affects power quality due to its nonlinearity, since solar generation plants and wind power generators must be connected to the grid through high-power static PWM converters. The non uniform nature of power generation directly affects voltage regulation and creates voltage distortion in power systems. This new scenario in power distribution systems will require more sophisticated compensation techniques. An active power filter implemented with a four-leg voltage-source inverter using a predictive control scheme is presented. This project presents the mathematical model of the 4L-VSI and the principles of operation of the proposed predictive control scheme, including the design procedure.

KEYWORDS: Voltage source inverter, Power quality, PWM converter, Predictive control scheme,

I. INTRODUCTION

Renewable generation affects power quality due to its nonlinearity, since solar generation plants and wind power generators must be connected to the grid through high-power static PWM converters [1]. The non uniform nature of power generation directly affects voltage regulation and creates voltage distortion in power systems. This new scenario in power distribution systems will require more sophisticated compensation techniques. Although active power filters implemented with three-phase four-leg voltage-source inverters (4L-VSI) have already been presented in the technical literature [2]–[6], the primary contribution of this paper is a predictive control algorithm designed and implemented specifically for this application. Traditionally, active power filters have been controlled using pre tuned controllers, such as PI-type or adaptive, for the current as well as for the dc-voltage loops [7], [8].

PI controllers must be designed Based on the equivalent linear model, while predictive controllers use the nonlinear model, which is closer to real operating conditions. So far, implementations of predictive control in power convertershave been used mainly in induction motor drives [9]–[16]. In the case of motor drive applications, predictive controlrepresents a very intuitive control scheme that handles multivariable characteristics, simplifies the treatment of dead-time compensations, and permits pulse-width modulator replacement. However, these kinds of applications present disadvantages related to oscillations and instability created from unknown load parameters [15].

These power quality concerns made the power engineers to think about the devices which reduces the harmonics in the supply line [E,F].Such devices are known as active power filter/power conditioners which are capable of current/voltage harmonic compensation. Active power filters are classified into shunt , series and hybrid active power filters which can deal with various power quality issues [illegible] Nowadays power quality issues in single One advantage of the proposed algorithm is that it fits well in active power filter applications, since the power converter output parameters are well known [17]. These output parameters are obtained from the converter output ripple filter and the power system equivalent impedance. The converter output ripple filter is part of the active power filter design and the power system impedance is obtained from well-known standard procedures [18], [19]. In the case of unknown system impedance parameters, an estimation method can be used to derive an accurate *R–L*

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

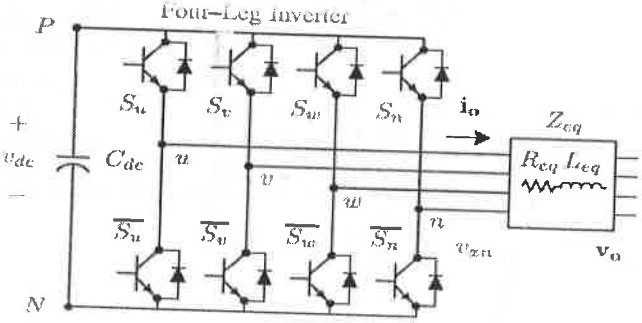


Fig. 3 Two-level four-leg PWM-VSI topology.

$$V_{xn} = S_x - S_n V_{dc}$$
$$X = u,v,w,n. \qquad (1)$$

The mathematical model of the filter derived from the equivalent circuit shown in Fig. 2 is

$$V_o = V_{xn} - R_{eq} i_o - L_{eq} \frac{di_o}{dt} (2)$$

where*R*eq and *L*eq are the 4L-VSI output parameters expressed as Thevenin impedances at the converter output terminals *Z*eq . Therefore, the Thevenin equivalent impedance is determined by a series connection of the ripple filter impedance *Zf*and a parallel arrangement between the system equivalent impedance *Zs*and the load impedance *ZL*

$$Z_{eq} = \frac{Z_s Z_L}{Z_s + Z_L} + Z_f \approx Z_L + Z_f \quad (3)$$

For this model, it is assumed that *ZL* _ *Zs*, that the resistive part of the system's equivalent impedance is neglected, and that the series reactance is in the range of 3–7% p.u., which is an acceptable approximation of the real system. Finally, in (2) *R*eq = *Rf*and*L*eq = *Ls*+ *Lf* .

III. DIGITAL PREDICTIVE CURRENT CONTROL

The block diagram of the proposed digital predictive current control scheme is shown in Fig. 4. This control scheme is basically an optimization algorithm and, therefore, it has to be implemented in a microprocessor. Consequently, the analysis has to be developed using discrete mathematics in order to consider additional restrictions such as time delays and approximations [10], [22]–[27]. The main characteristic of predictive control is the use of the system model to predict the future behavior of the variables to be controlled. The controller uses this information to select the optimum switching state that will be applied to the power converter, according to predefined optimization criteria. The predictive control algorithm is easy to implement and to understand, and it can be implemented with three main blocks, as shown in Fig. 4.

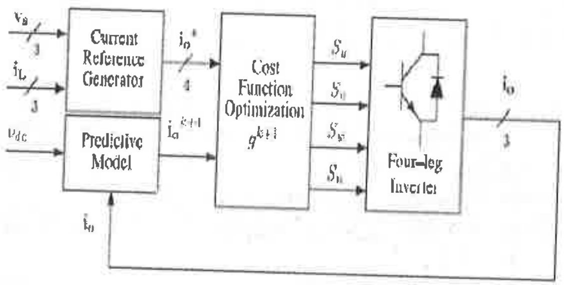


Fig. 4 Proposed predictive digital current control block diagram.

DOI:10.15662/IJAREEIE.2015.0506001

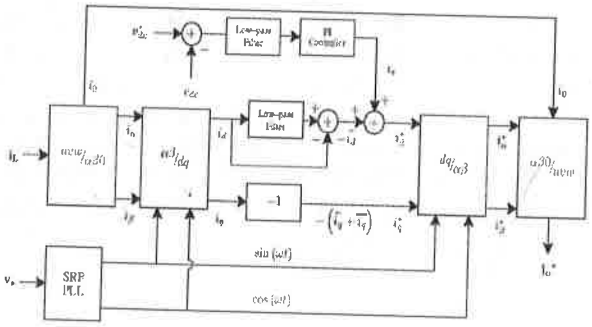


Fig. 5 *dq*-based current reference generator block diagram.

The sin(*wt*) and cos(*wt*) synchronized reference signals are obtained from a synchronous reference frame (SRF) PLL [29]. The SRF-PLL generates a pure sinusoidal waveform even when the system voltage is severely distorted. Tracking errors are eliminated, since SRF-PLLs are designed to avoid phase voltage unbalancing, harmonics (i.e., less than 5% and 3% in fifth and seventh, respectively), and offset caused by the nonlinear load conditions and measurement errors [30]. Equation (8) shows the relationship between the real currents *iLx*(*t*) (*x* = *u, v,w*) and the associated *dq*components (*id* and *iq*)

$$\begin{bmatrix} i_d \\ i_q \end{bmatrix} = \sqrt{\frac{2}{3}}\begin{bmatrix} \sin wt & \cos wt \\ -\cos wt & \sin wt \end{bmatrix}\begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix}\begin{bmatrix} i_{Lu} \\ i_{Lv} \\ i_{Lw} \end{bmatrix} \qquad (8)$$

Alow-pass filter (LFP) extracts the dc component of the phase currents *id* to generate the harmonic reference components −_*id* . The reactive reference components of the phase-currents are obtained by phase-shifting the corresponding ac and dc components of i_q by 180. In order to keep the dc-voltage constant, the amplitude of the converter reference current must be modified by adding an active power reference signal *ie*with the *d*-component, as will be explained in Section IV-A. The resulting signals *i∗d*and*i∗q*are transformed back to a three-phase system by applying the inverse Park and Clark transformation, as shown in (9). The cutoff frequency of the LPF used in this project is 20 Hz The current that flows through the neutral of the load is compensated by injecting the same instantaneous value obtainedthird-order harmonic content, and system current imbalance (withrespect to positive sequence of the system current, *is,*1).from the phase-currents, phase-shifted by 180◦, as shown next.

$$\begin{bmatrix} i_{ou}^{*} \\ i_{ov}^{*} \\ i_{ow}^{*} \end{bmatrix} = \sqrt{\frac{2}{3}}\begin{bmatrix} \frac{1}{\sqrt{2}} & 1 & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{1}{\sqrt{2}} & -\frac{1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sin wt & -\cos wt \\ 0 & \cos wt & \sin wt \end{bmatrix}\begin{bmatrix} i_o \\ i_d^{*} \\ i_q^{*} \end{bmatrix} \qquad (9)$$

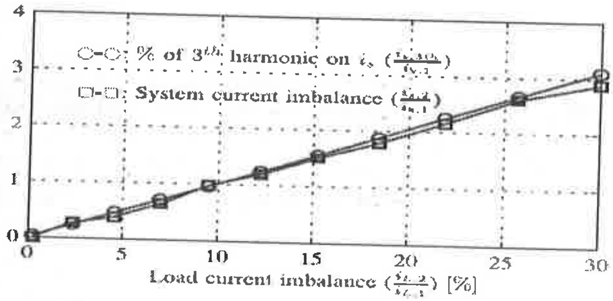


Fig. 6 Relationship between permissible unbalance load currents, the corresponding

DOI:10.15662/IJAREEIE.2015.0506001

V. SIMULATION RESULTS

A simulation model for the three-phase four-leg PWM converter with the parameters shown in Table I has been developed usingMATLAB-Simulink. The objective is to verify the current harmonic compensation effectiveness of the proposed control scheme under different operating conditions. A six-pulse rectifier was used as a nonlinear load. The proposed predictive control algorithm was programmed using an S-function block that allows simulation of a discrete model that can be easily implemented in a real-time interface (RTI) on the dSPACE DS1103 R&D control board. Simulations were performed considering a 20 [μs] of sample time. In the simulated results shown in Fig. 8, the active filter starts to compensate at $t = t1$. At this time, the active power filter injects an output current *iout*to compensate current harmonic components, current unbalanced, and neutral current simultaneously.

variable	Description	Value
V_s	Source voltage	55[v]
f	System frequency	50[Hz]
V_{dc}	dc-voltage	162[v]
C_{dc}	dc capacitor	2200[μF] (2.0 pu)
L_f	Filter inductor	5.0[mH] (0.5 pu)
R_f	Internal resistance within L_f	0.6[Ω]
T_s	Sampling time	20[μs]
T_e	Execution time	16[μs]

Table. 1: Specification paramerers.

During compensation, the system currents is show sinusoidal waveform, with low total harmonic distortion (THD = 3.93%). At $t = t2$, a three-phase balanced load step change is generated from 0.6 to 1.0 p.u.

The compensated system currents remain sinusoidal despite the change in the load current magnitude. Finally, at $t = t3$, a single-phase load step change is introduced in phase *u* from 1.0 to 1.3 p.u., which is equivalent to an 11% current imbalance.As expected on the load side, a neutral current flows through the neutral conductor (*iLn*), but on the source side, no neutral current is observed (*isn*). Simulated results show that the proposed control scheme effectively eliminates unbalanced currents.

DOI:10.15662/IJAREEIE.2015.0506001

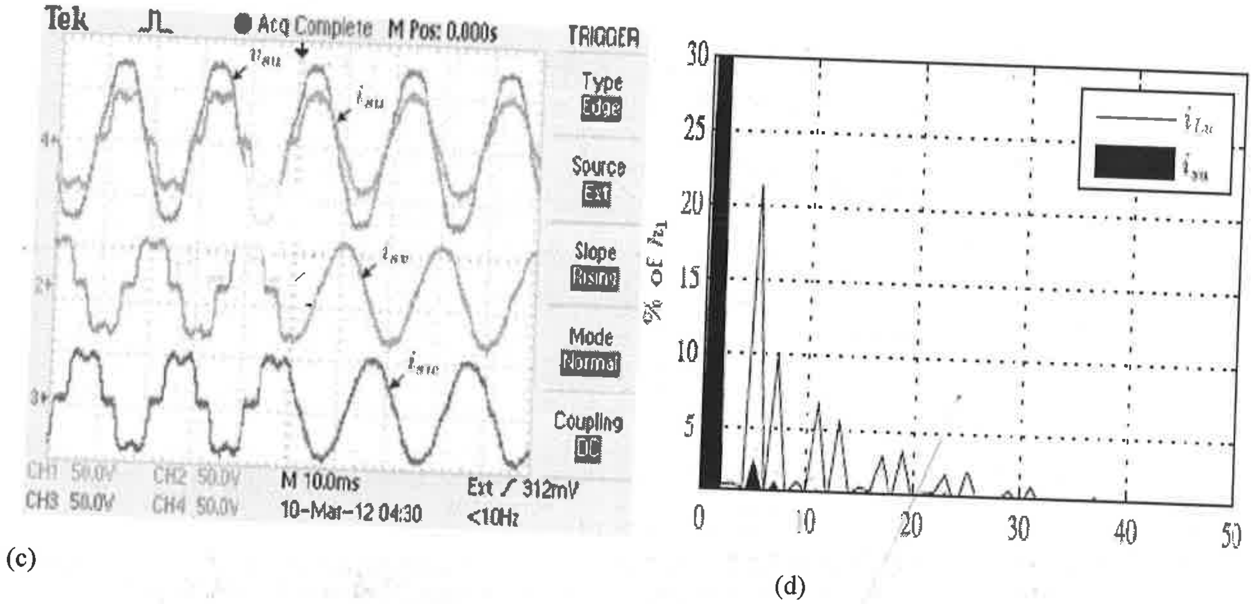


Figure. 9. Experimental transient response after APF connection. (a) Load Current *iLu*, active power filter current *iou*, (b)dc-voltage converter vdc, and system current *isu*. Associated frequency spectrum. (c) Voltage and system waveforms, *vsu*and*isu* ,*isv* , *isw*. (d) Current reference signals *i***ou*, and active power filter current *iou*(tracking characteristic).

Fig. 9 showsthe transient response of the compensation scheme. Fig. 9(a) shows that the line current becomes sinusoidal when the active power filter starts compensation, and the dc-voltage behaves as expected. Experimental results shown in Fig. 9(b) indicate that the total harmonic distortion of the line current (THD*i*) is reduced from 27.09% to 4.54%. This is a consequence of the good tracking characteristic of the current references, as shown in Fig. 9(d).

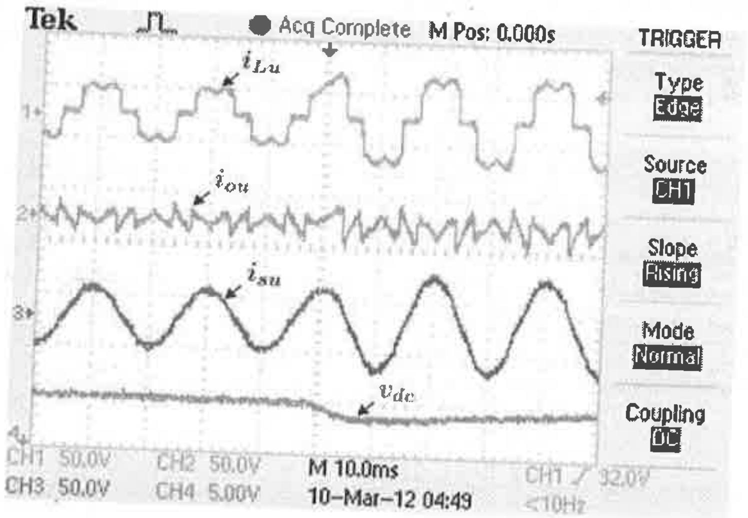


Figure. 10. Experimental results for step load change (0.6 to 1.0 p.u.). Load Current *iLu*, active power filter current *iou*, system current *isu*, and dc-voltage converter *v*dc .

International Journal of
Scientific Engineering and Technology Research

ISSN 2319-8885
Vol.04,Issue.36,
September-2015,
Pages:7899-7905

www.ijsetr.com

A New Hybrid Power Conditioner for Suppressing Harmonics and Neutral Line Current in Three Phase Four Wire Distribution Power System

AMOL PADMAKAR CHOPADE[1], J. SHANKAR[2]

[1]PG Scholar, Dept of EEE(EPS), Annamacharya Institute of Technology and Sciences, Hyderabad, India,
E-mail: amolpchopade@gmail.com.

[2]Assistant Professor, Dept of EEE, Annamacharya Institute of Technology and Sciences, Hyderabad, India.

Abstract: In this paper, a new hybrid power conditioner is proposed for suppressing harmonic currents and neutral-line current in three-phase four-wire distribution power systems. The proposed hybrid power conditioner is composed of a neutral-line current attenuator and a hybrid power filter. The hybrid power filter, configured by a three-phase power converter and a three-phase tuned power filter, is utilized to filter the nonzero-sequence harmonic currents in the three-phase four-wire distribution power system. The three-phase power converter is connected to the inductors of the three-phase tuned power filter in parallel, and its power rating can thus be reduced effectively. The tuned frequency of the three-phase tuned power filter is set at the fifth harmonic frequency. The neutral- line current suppressor is connected between the power capacitors of the three-phase tuned power filter and the neutral line to suppress the neutral-line current in the three-phase four-wire distribution power system. With the major fundamental voltage of the utility dropping across the power capacitors of the three-phase tuned power filter, the power rating of the neutral-line current suppressor can thus be reduced. Hence, the proposed hybrid power conditioner can effectively reduce the power rating of passive and active elements. A hardware prototype is developed to verify the performance of the proposed hybrid power conditioner. Experimental results show that the proposed hybrid power conditioner achieves expected performance.

Keywords: Dynamic-Implications of Technology, Technology Social Factors, Privacy.

I. INTRODUCTION

Three-Phase four-wire distribution power systems been widely applied in office buildings and manufacturing- office buildings to supply single-phase or three-phase loads. The third harmonic is very serious in single-phase nonlinear loads. The third-order harmonic current of each phase is synchronous and regarded as the zero-sequence current. Therefore, the zero-sequence currents of each phase are summed up and flow into the neutral line of three-phase four-wire distribution power systems. Furthermore, single-phase loads may result in serious load unbalance, and the unbalanced load current also flows into the neutral line of the three-phase four-wire distribution power systems. In many applications, the neutral-line current will exceed the phase currents. Excessive neutral-line current may cause accidents due to overload of the neutral line. Moreover, it will lead to fluctuation in ground voltage of the load, which may influence the operation of precision equipment. Hence, the major problems of three-phase four-wire distribution power systems are harmonic currents and neutral-line current [1], [2]. The zig-zag transformer, connected to the load in parallel, has been employed to attenuate the neutral-line current [1], [3], [4]. However, the attenuation of neutral-line current is dependent on the ratio between the impedance of the utility system and the zig-zag transformer. Furthermore, the zig-zag transformer also has a low impedance path for zero-sequence voltage of the unbalanced utility, which will further cause a significant neutralline current [4]. A single-phase power converter can be combined with the zig-zag transformer to advance the performance of the neutral-line current suppression [5], [6].

The single-phase power converter is inserted at the neutral line between the load and the utility, thus causing fluctuation in the ground voltage of the load. A neutral-current suppression scheme, configured by a -Y transformer and a single-phase power converter connected in series, is connected to the load in parallel to suppress the neutral- line current [7]. The neutral line of the load is directly connected to that of the utility, and the fluctuation in ground voltage of the load can thus be avoided. A series of active power filters connected to the neutral line between the utility and the load can suppress the neutral-line current, thus eliminating the need of the transformer for a zero current path [8]. However, there is fluctuation in ground voltage of the load because the neutral lines of the load and utility are separated. Conventionally, passive power filters have been employed to solve the problems of harmonic currents and neutral-line current in three-phase four-wire distribution power systems. Although passive power filters have the advantage of low hardware cost, their performance is often significantly affected by the system impedance. Furthermore, salient problems, including large volume, parallel resonance, and

A New Bidirectional Intelligent Semiconductor Transformer for Smart Grid Applications

T.Anjakumar
M-tech Student Scholar
Department of Electrical & Electronics Engineering,
Annamacharya Institute of Technology and Sciences,
Piglipur(v); Batasingaram(post),Hayathnagar(Mandal), Ranga reddy (Dt); T.G, India.

J.Shankar
Assistant Professor
Department of Electrical & Electronics Engineering,
Annamacharya Institute of Technology and Sciences,
Pigilipur(v), Batasingaram(post); Hayathnagar(Mandal) Ranga reddy(Dt); T.G, India.

Abstract—**This paper proposes a new bidirectional intelligent semiconductor transformer (BIST) for the smart distribution system and smart grid. The proposed BIST consists of high-voltage high-frequency ac/dc converter, bidirectional low-voltage dc/dc converter, and hybrid-switching dc/ac inverter. It features 1) input to- output isolation with a high-frequency transformer; 2) bidirectional power flow; 3) small size and light weight; 4) capability of compensating voltage sag and/or swell; and 5) realization of three-phase structure based on single-phase module. The operational feasibility of proposed transformer was verified not only by computer simulation with PSCAD/EMTDC software but also by a hardware prototype with rating of 1.9 kV/127 V, 2 kVA, allowing a three-phase transformer of 3.3 kV/220 V, 6 kVA with three-phase construction.**

Index Terms—**Bidirectional dc/ac converter, bidirectional intelligent semiconductor transformer (BIST), high-voltage ac/dc rectifier, hybrid-switching, PSCAD/EMTDC.**

I. Introduction

Conventional transformer composed of coil and iron core can change only the magnitude of the ac voltage and the quality of supplying power is totally dependent on that of the input power. So, it cannot be applicable for the smart grid, in which the magnitude and frequency of the operation voltage are various and high-quality power is required. Intelligent semiconductor transformer or solid-state transformer was proposed by EPRI to replace the conventional transformer in railway systems and substations, in which light weight is mandatorily required [1]. Recently, EPRI has reported 100 kVA single-phase semiconductor transformer named intelligent universal transformer for distribution automation [2]. Intelligent semiconductor transformer can easily offer small size and light weight because it operates at much higher frequency with reduction of the magnetic component. It can supply not only the dc power, but also high-quality ac power to the customer by compensating the voltage sag, swell, and harmonics. So, it can be utilized for implementing the smart distribution system and the micro grid [3]–[5]. Various kinds of intelligent semiconductor transformers were already proposed. However, since the power flow in these transformers is unidirectional, it is not properly applicable for the dc distribution and micro grid [1], [2], [6]–[10]. One can find some studies on the semiconductor transformer topologies with bidirectional power flow capability [11]–[23]. In [11] and [12], bidirectional power flow can be achieved but the power factor is not controlled. The topology in [13] can compensate sag/swell voltage; however, it employs heavy and bulky line- frequency transformer for isolation. The semiconductor transformer in [14] has not only the bidirectional power flow functions but also voltage sag compensation where high-frequency dc/dc power conversion is employed. The circuit configuration in [14], however, shows too many active switching device counts, at least 18 IGBTs for implementing single phase module. In [15]–[23], three-stage structure comprised of ac/dc converter, dual-active-bridge dc/dc converter, and inverter. These topologies provide power factor correction and reactive power compensation, but they suffer from heavy turn-off loss in dc/dc stage and complex control for voltage balancing.

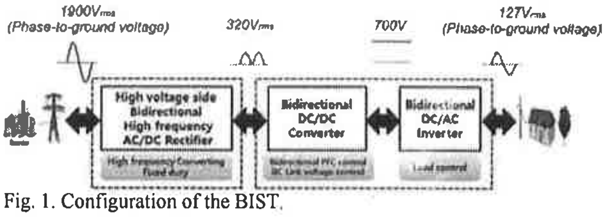

1900V rms
(Phase-to-ground voltage)
320V rms
700V
127V rms
(Phase-to-ground voltage)
High voltage side Bidirectional High frequency AC/DC Rectifier
High frequency Converting Fixed duty
Bidirectional DC/DC Converter
Bidirectional PFC control DC Link voltage control
Bidirectional DC/AC Inverter
Load control

Fig. 1. Configuration of the BIST.

This paper proposes a new bidirectional intelligent semiconductor transformer (BIST) for the smart distribution system and micro grid. The proposed BIST consists of high-voltage part and low-voltage part, whose configuration is shown in Fig. 1.

The high-voltage part is composed of several half-bridge ac/dc converters connected in series through high frequency transformers to cope with high input voltage, while the low-voltage part is composed of bidirectional half-bridge dc/dc converter and dc/ac PWM inverter. In the prototype BIST, the input voltage on the high-voltage side is 1900 V and the output voltage on the low-voltage side is 127 V, in which the primary and secondary dc-link voltages are 320 V and 700 V, respectively. A three-phase 3.3 kV/220 V transformer can be built using three units of 1.9 kV/127 V single-phase module.

ANNAMACHARYA INSTITUTE OF TECHNOLOGY & SCIENCES

Piglipur, Batasingaram, Hayathnagar (M), Hyderabad, R.R.Dist-501512

3.3.2 Number of research papers per teachers in the Journals notified on UGC website during the last five years (10)

Title of paper	Name of the author/s	Department of the teacher	Name of journal	Year of publication	ISSN number	Link to the recognition in UGC enlisted Journal / Digital Object Identifier (doi) Number
1.“Two Line Resolution of Defocusing with Shading of theApertures	T.Kiran Kumar	H&S	International Journal of Innovative Research & Studies (ijirs)	2014	2319-9725	http://ijirs.in/
2. “Shaping of the Aperture with coma in Two Line Resolution”	T.Kiran Kumar	H&S	International Journal of Research in Pure and Applied Physics (ijrpap)	2014	0973-1776	http://www.ripublication.com/ijaer.htm
3. “Apodisation in Two – Line Resolution by Shaping of the Aperture”	T.Kiran Kumar	H&S	ISST Journal of Applied Physics	2014	0976-7363	www.isst.org.in
4. “Point Spread Functions Of Aberrated Optical Systems With Annular Apertures”	T.Kiran Kumar	H&S	International Journal of Scientific Research and Engineering Studies (ijsres)	2014	2349-8862	http://www.ijsres.com/
5. “Two line resolution of apodised optical system with circular apertures in the presence of primary spherical aberration, defocusing and coma”	T.Kiran Kumar	H&S	International Journal of Engineering Research and Management(IJERM)	2014	2349-2058	https://www.ijerm.com/
1.Aperture shaping for Two Line resolution in the apodised variable coherence in The presence of Primary Spherical Aberration	T.Kiran Kumar	H&S	International Journal of Research and Development Organization (IJRDO)	2015	2456-1843	https://www.ijrdo.org/
2. Shaded apertures and Bartlet Apertures for Two Line Resolutions on aperture shaping	T.Kiran Kumar	H&S	Journal of The International Association of Advanced Technology and Science (JIAATS)	2015	ISSN-2095-1563	http://www.jiaats.com/

PRINCIPAL
Annamacharya Institute of
Technology & Sciences
Piglipur (V), Batasingaram (Post),
Abdullapurmet (M), R.R.Dist.HYD-501 512.

3. Two Line Resolutions with circular apertures and Band pass apertures in the presence of defocusing aberration	T.Kiran Kumar	H&S	EPH International Journal of Science and Engineering	2015	ISSN: 2394 - 3785	https://www.ephjournal.com/
4. Two Point Resolution on aperture shaping with Rayleigh and Sparrow Limits	T.Kiran Kumar	H&S	International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)	2015	ISSN:2394-3777	https://ijartet.com/
5.A Study on Behavioral Intentions of Teaching Faculties in Engineering Colleges Towards Research And Teaching.	K.Ramesh Gowd	H&S	Abhinav International Monthly Refereed Journal of Research in Management & Technology	2015	ISSN- 23200073	www.abhinavjournal.com
6.“Syntheis characterization and biological activity of some new of novel 1,3,4-oxadiazole bearing thiophene-2- carboxylic acid”	Sunder Kumar Kolli	H&S	World Journal of Pharmacy and Pharmaceutical Sciences,	2016	ISSN 2278 – 4357	[DOI : 10.20959/wjpps20165-6802]
7.“Musi River pollution effect on human anthropogenic activities”	Sunder Kumar Kolli	H&S	Research Journal of Chemical Sciences	2016	ISSN: 2231-606X	https://www.openacessjournal.com
1. Spectroscopic Evidence for Steric Enhancement of Resonance in Rhodium (III) Chloride Conplexes.	Sunder Kumar Kolli	H&S	World Journal of Pharmacy and Pharmaceutical Sciences	2017	ISSN 2278 – 4357	https://www.wjpps.com/
2. Superconductors in the field of communication and study various electric properties.	Sunder Kumar Kolli	H&S	International Journal of Science and Research (IJSR)	2017	ISSN 2319-7064	https://www.ijsr.net/
3.Climate change effects on sustainable scenario in India.	Sunder Kumar Kolli	H&S	BS publications	2017	ISBN 978-93-86819-94-9	https://www.bspublications.net
4.Ultrasound-Assisted 3-Component Reaction in Acetic Acid Alone: Catalyst/Promoter/Ligand Free Synthesis of Bioactive Pyrazolo[1,5-a]pyrimidines.	Sunder Kumar Kolli	H&S	Letters in Drug Design & Discovery	2017	ISSN-1570-1808	DOI: https://doi.org/10.2174/1570180814666170126120408

WORLD JOURNAL OF PHARMACEUTICAL RESEARCH
SJIF Impact Factor 7.523
Volume 6, Issue 10, 901-923. Research Article ISSN 2277– 7105

SPECTROSCOPIC EVIDENCE FOR STERIC ENHANCEMENT OF RESONANCE IN RHODIUM (III) CHLORIDE COMPLEXES

Penke Vijaya Babu[1], Sunder Kumar Kolli[2] and Manam Sreenivasarao[3]*

[1]Department of Organic Chemistry, The Weizmann Institute of Science, Rehovot, Israel, 76100.

[2]Department of Chemistry, Annamacharya Institute of Technology Sciences, Hyderabad, T.S. India.501512.

[3]Department of Chemistry, Chundi Ranganayakulu college, Chilakaluripet, Guntur, A.P. 522616, India.

Article Received on
10 July 2017,
Revised on 30 July 2017,
Accepted on 20 Aug 2017
DOI: 10.20959/wjpr201710-9393

***Corresponding Author**
Manam Sreenivasarao
Department of Chemistry,
Chundi Ranganayakulu
college, Chilakaluripet,
Guntur, A.P. 522616,
India.

ABSTRACT

The assignments reveal that the carboxyl substituted complexes are isolated as fac- isomers and all the remaining seven complexes are having mer-octahedral configuration. Using these sulphides as ligands Rhodium (III) chloride complexes have been synthesized and analyzed. The infrared spectra of these complexes have been recorded in the region 600– 200 cm^{-1}. Generally the electron releasing group in Benzene ring decreases the ν(Rh-S) stretching frequency value from the present one where as the electron withdrawing group increases the ν(Rh-S) stretching frequency value form its parent one. Otherwise the electron donor exerts a weakening effect on the Rh-S band and the σ-electron acceptor removes from the metal. The appearance of –OH bond in the complexes of carboxylic substituted ligands shows the absence of intra molecular hydrogen bonding between –OH of carboxylic group and chlorine. The carboxyl frequency of meta and para carboxylic groups gets increased when compared with the ligands. But in the case of higher stability of the meta and para isomers. In recording the electronic spectra the same solvent was used both for the complex and ligand. Generally the bathochromic shift is observed when there is an electron releasing group present in the benzene ring and the presence of electron withdrawing group or steric inhibition results in a hypsochromic shift. The complexes involving p-NO_2 and p-COOH substituted phenyl methyl sulphides have the absorption bands almost identical with those of ligands. When an electron donor is present in

WORLD JOURNAL OF PHARMACY AND PHARMACEUTICAL SCIENCES

SJIF Impact Factor 6.041

Volume 5, Issue 5, 1800-1825 Research Article ISSN 2278 – 4357

SYNTHESIS, CHARACTERIZATION AND BIOLOGICAL ACTIVITY OF SOME NEW OF NOVEL 1,3,4-OXADIAZOLE BEARING THIOPHEN-2-CARBOXYLIC ACID

Sunder Kumar Kolli*

Department of Chemistry, Annamacharya Institute of Technology Sciences Piglipur,
Bhatasingaram, Hayath Nagar, Hyderabad, R.R. Dist. 501512.

Article Received on
22 March 2016,

Revised on 12 April 2016,
Accepted on 02 May 2016

DOI: 10.20959/wjpps20165-6802

***Corresponding Author**
Dr. Sunder Kumar Kolli
Department of Chemistry,
Annamacharya Institute of
Technology Sciences
Piglipur, Bhatasingaram,
Hayath Nagar, Hyderabad,
R.R. Dist. 501512.

ABSTRACT

Oxadiazole compounds are various active pharma ingredients which exhibit different potent biological activities including antibacterial, antifungal and anticancer activities. Now our research concentrates the synthesis of different oxadiazole derivatives. Different carboxylic acid derivatives treated with ethanol in the presence of Conc.acids to give the intermediates then further its react with hydrazine hydrate followed by various carboxylic acida in the presence of $POCl_3$ to give oxadiazole derivatives. All the newly synthesized compounds were synthesized compounds were screened for their antibacterial and antifungal studies.

KEYWORDS: Thiophehe-2-Carboxylic acid, $POCl_3$, antibacterial, antifungal activity and 2,5- disubstituted-1,3,4-oxadiazoles.

INTRODUCTION

Oxadiazole derivatives which are widely used in pharmacy, dyestuff production, scintillation technology, production of thermo-stable polymers and in other fields of science and technology. 2,5-substituted 1,3,4-oxadiazoles were synthesized in the end of the last century, but there have been no reviews on this subject, embracing the literature up to 1961 and including nearly all the most important material on the preparation and chemical properties of 2,5-disubstituted 1,3,4-oxadiazoles.

The presence of hetero atoms such as oxygen, nitrogen and sulfur in five, six-membered and fused ring systems are very important in the synthesis of several natural products. The natural products associated with imidazole, thiazole, pyrazole, pyrazine, furan, pyrrole, 1,3,4-

Superconductors in the Field of Communication and Study Various Electric Properties

Dr. Sunder Kumar Kolli[1], Sukumar Velpula[2], Kiran Kumar Yenubari[3]

[1]Department of Chemistry, Annamacharya Institute of Technology & Sciences, Hyderabad

[2]Department of Physics, Arjun College of Technology & Sciences, Hyderabad

[3]Department of Physics, Sreyas Institute of Engineering & Technology, Hyderabad

Abstract: ***Zero resistance and high current density have a major impact on electric power transmission and also enable much smaller or more powerful magnets for motors, generators, energy storage, medical equipment and industrial separations. Low resistance at high frequencies and extremely low signal dispersion are key aspects in microwave components, communications technology and several military applications. Low resistance at higher frequencies also reduces substantially the challenges inherent to miniaturization brought about by resistive, or I^2R, heating. The high sensitivity of superconductors to magnetic field provides a unique sensing capability, in many cases 1000x superior to today's best conventional measurement technology. Superconductors in the field of communication and study of various the electric properties such as Conductors, Semiconductors, Insulator and Superconductors. Superconductors specifications graphs in communication feature scope- Superconductor in the field of communication towers such that loss of electrical radiation to protect our environment.***

Keywords: Zero resistance, Superconductor, Magnetic field and Intermetallic compounds

1. Introduction

In 1911, H. K. Onnes, a Dutch physicist, discovered superconductivity by cooling mercury metal to extremely low temperature and observing that the metal exhibited zero resistance to electric current. Prior to 1973 many other metals and metal alloys were found to be superconductors at temperatures below 23.2K. These became known as Low Temperature Superconductor (LTS) materials. Since the 1960s a Niobium-Titanium (Ni-Ti) alloy has been the material of choice for commercial superconducting magnets. More recently, a brittle Niobium-Tin intermetallic material has emerged as an excellent alternative to achieve even higher magnetic field strength.[1] In 1986, J. G. Bednorz and K. A. Müller discovered oxide based ceramic materials that demonstrated superconducting properties as high as 35K. This was quickly followed in early 1997 by the announcement by C. W. Chu of a cuprate superconductor functioning above 77K, the boiling point of liquid nitrogen. Since then, extensive research worldwide has uncovered many more oxide based superconductors with potential manufacturability benefits and critical temperatures as high as 135K. A superconducting material with a critical temperature above 23.2K is known as a High Temperature Superconductor (HTS), despite the continuing need for cryogenic refrigeration for any application.

Arising the scope of studying superconductor in the field of communication first discuss about superconductivity of different materials. Super conductors cope in transmission of signals in terms of electromagnetic effects such as meissner effect, ac josphen, dc josphen effect, zero resistance. The experiment carried about past decades for the development of communication transmission and receiving of signals of superconducting probes or materials. Transmission of signal such as antennas, radar, transmitter and receiver to travel signals for longer distance. Building block of superconducting materials in satellite communication in ISRO, NASA etc. growth and development of superconducting materials in satellite, electromagnetic waves in the field of communication.

2. Unique Properties

- Zero resistance to direct current
- Extremely high current carrying density
- Extremely low resistance at high frequencies
- Extremely low signal dispersion
- High sensitivity to magnetic field
- Exclusion of externally applied magnetic field
- Rapid single flux quantum transfer
- Close to speed of light signal transmission

Challenges
- Cost
- Refrigeration
- Reliability
- Acceptance

Transportation: The rapid and efficient movement of people and goods, by land and by sea, poses important logistical, environmental, land use and other challenges. Superconductors are enabling a new generation of transport technologies including ship propulsion systems, **Industry:** magnetically levitated trains and railway traction transformers.

Medicine: Advances in HTS promise more compact and less costly Magnetic Resonance Imaging (MRI) systems with superior imaging capabilities. In addition, Magneto-Encephalography (MEG), Magnetic Source Imaging (MSI) and Magneto-Cardiology (MCG) enable non-invasive diagnosis of brain and heart functionality.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

Assessment of Ground Water Quality in Medchal-Malkajgiri District

Srisailam Gogula[1], David Wilson Narisinga[2], Mohammed Vaseem[3], Sunder Kumar Kolli[4]
[1, 2, 3] Department of Chemistry,Govt. City College, Hyderabad, Telangana.
[4]Department of Chemistry, Annamacharya Institute of Technology Sciences, Hyderabad

Abstract: Water is the most important substance in our daily life. Without water, life would not have beenpossible. The magnitude of water problem is increased due to poor drainage system, unplanned industries, increase of pollution, influxes of people from rural areas and other human activities. Due to rapid increase in population, urbanization and industrialization in Hyderabad have resulted the drastic increase in water pollution, which is one of the largest and smart city in India. In this study the ground water samples are collected in different seasons i.e., pre-monsoon and post-monsoon in the year 2016 for analysis from various places of Rangareddy district (Medchal) in Telangana State. The Physico-chemical parameters such as pH, total dissolved solids (TDS), chloride (Cl^-), fluoride (F^-), nitrate(NO_3^-), Sulphate(SO_4^{-2}), hardness ($CaCO_3$, $MgCO_3$), sodium(Na), potassium(K) are analyzed with different analytical methods used by technical instruments. The results were compared with standard values given by World Health Organization (WHO). The present study revealed that the parameters of water which is too higher than the standard limits.

Keywords: Physicochemical parameters, Water pollution, Total Hardness, Fluoride and Nitrate.

I. INTRODUCTION

India is the biggest developing country having 1.3 billion population, it needs to provide more facilities in various sectors for the peoples' sustainability. It is developing in agriculture in rural level, organizations and industries at urban areas. Urbanization is more because growth rate is more at urban areas like Bengaluru, Hyderabad, Amaravati etc. Hyderabad is situated at the banks of Musi river and it has great history, it is the capital city of Telangana state and Andhra Pradesh. Hyderabad city has good climate and it provides many resources for the people to settle their career in various fields, so many are coming to Hyderabad city from various places of India mostly from Telangana and Andhra Pradesh states and the Hyderabad city is ranked as the best city in India in living standards by Mercer's Quality of Life Index ranking in 2017, thus many choose the city as permanent place as it gives multi careers. It iswell developed in various sectors like IT, Pharma, educational etc. and it has major industrial areas in and around the city.Medchal is also one of the industrial area in city, which is carved out of erstwhile Ranga reddy district, now it is re-organized as a district by the Telangana government in 2016 named as Medchal-Malkajgiri district.Generally due to urbanization[1-5] and industrialization air and water gets pollution, especially water quality reduces, which becomes more dangerous[6]. Water is essential for human life and needs. Natural and ground water is more affected at such areas like Medchal.

II. MATERIALS AND METHODS

Samples were collected from various selected sites of Medchal-Malkajgiri district of Hyderabad city in both pre- monsoon and post monsoon seasons, general analytical methods were used to assess the water quality of samples such as Electrical Conductivity meter, pH meter, Ion meter, UV Spectrophotometer, Nephelo meter and Flame photometer was used to test the water quality parameters in both the seasons. While collecting the samples location of sampling area noted and numbering is given for the sampling bottles from 1 to 20, Latitude and Longitudes are also noted and mentioned below in the table-1.

Table1: Latitude and Longitudes values of sampling sites

S. No	Name of the site	Latitude & Longitude	S. No	Name of the site	Latitude & Longitude
1	Medchal	17.6305° N, 78.4842° E	11	Kukatpally	17.4849° N, 78.3996° E
2	Jeedimetla	17.5172° N, 78.4612° E	12	Shapur	17.5394° N, 78.2675° E
3	Shamirpet	17.5895° N, 78.5706° E	13	Fathenagar	17.4573° N, 78.4517° E
4	Kapra	17.4888° N, 78.5718° E	14	Bowenpally	17.4765° N, 78.4884° E

International Journal of Advance Research in Science and Engineering
Volume No.07, Issue No.03, March 2018
www.ijarse.com

IJARSE
ISSN: 2319-8354

An Emphasis of Fluoride Effect on Human health and Treatment- Review

Srisailam Gogula[1], Mohammed Vaseem[1], S.Srilalitha[3], R. Indushree[4], Sunder Kumar Kolli[2]*

[1]Department of Chemistry, Govt. City College, Hyderabad, Telangana (India)
[2]Department of Chemistry, Annamacharya Institute of Technology & Sciences, Hyderabad (India)
[3]Department of Chemistry, Ashoka Institute of Engineering and Technology, Hyderabad (India)
[4]Department of Botany, P.E.S College of Science, Arts and Commerce, Mandy, Karnataka (India)

ABSTRACT

Water is the major medium of fluoride intake by humans. Fluoride in drinking water can be either beneficial or detrimental to health, depending on its concentration. The presence of fluoride in drinking water within permissible limits is beneficial in the calcification of dental enamel. According to the World Health Organization (WHO), the maximum acceptable concentration of fluoride is 1.5 mg/l, South Africa's acceptable limit is 0.75 mg/l, while India's permissible limit of fluoride in drinking water is 1 mg/l. Concentrations beyond these standards have shown dental and skeletal fluorosis and lesions of the endocrine glands, thyroid and liver. Fluoride stimulates bone formation and small concentrations have beneficial effects on the teeth by hardening the enamel and reducing the incidence of caries. Fluoride is a ubiquitous element present in earth's crust and is also being added to the environment anthropogenically. It is the most electronegative of all elements. Fluorine is found in the soil and the content of Fluorine in the lithosphere varies between 100 and 1500 g/ton. Fluoride has gained importance due to its dual influences on human beings. In lower concentrations, Fluoride is an essential nutrient which aids in the formation of bones, prevents tooth decay, whereas in higher concentrations it causes fluorosis, brittling of bones, curvature of bones, dwarfishness, mental derangements, cancer and in extreme cases even death. According to WHO standards, the Fluoride in drinking water should be within a range that slightly varies above and below 1 mg/L. In temperate regions, where water intake is low, Fluoride level up to 1.5 mg/L is acceptable. The Bureau of Indian Standards, BIS (IS-10500) has prescribed a desirable limit and permissible limit of Fluoride in drinking water as 1.0 and 1.5 mg/l respectively.

Keywords: Fluorosis, WHO, Concentration, human body, lithosphere and permissible limits.

I.INTRODUCTION

Water is one of the major elements essential for sustenance of all forms of life and is available in abundance in nature covering approximately three fourths of the surface of the earth. The chemical nature of water is one of the most important criteria that determines its usefulness for a specific need and as such not all the waters are fit for drinking; hence the problems of scarcity of drinking water. Over the year's groundwater has generally been

International Journal of Advance Research in Science and Engineering
Volume No.07, Issue No.03, March 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

Apodised Optical System In The Presence Of Shrink apertures For Two-Line Resolution

S.Radhika[1], Sukumar[2], P.Pothanna[3], T.Kiran Kumar[4]

[1]Assistant Professor, Department of Physics, Govt Degree College, Khairtabad, Hyd (India)
[2]Assistant Professor, Department of Physics, Sri Indu Engg. College, Hyderabad (India)
[3]Assistant Professor, Department of Physics, AITS Hyd (India)
[4]Assistant Professor, Department of Physics, AITS Hyd (India)

ABSTRACT

The optical system having an apodised circular aperture suffering from the effects of spherical aberrations. Analytical investigations have been carried out on the intensity distribution of the Two-Line objects in the image plane by primary spherical aberration. Studies were also made on the imaging characteristics of the optical systems subjected to spherical aberration. The individual influence of the apodisation, spherical aberration on the Two-Line Resolution have been examined.

I.INTRODUCTION

Generally the optical systems can be classified into three different categories such as: ideal optical systems, perfect optical systems and real optical systems. An ideal optical system is that in which both diffraction and aberrations are completely eliminated. Hence, an ideal optical system gives a point image of a point object and straight-line image of a straight line. That means the image and objects should be in a projective relation to each other. In the investigations on the general resolution problem in optical systems ARSAC [1956] has discussed the problem with Fourier integral theory.

II. MATHEMATICAL FORMULATION

When the optical system is apodised, each point gives rise to a diffraction image whose normalized amplitude response to unit amplitude in the object point. f(r) is the chosen amplitude filter. In the present study the following filters are employed: $f(r_1)= \cos(\pi\beta r)$ Hanning Amplitude Filter , $f(r_2) = (1-\beta r)$ – Bartlett Filter, $f(r_3) = (1-\beta r^2)$ – Shaded Aperture Filter and $f(r_4) = \sin(\pi\beta r) / \pi\beta r$ – Lancoz Filter and c is the intensity ratio between the Two Lines and Z_0 is distance of separation the integral 0 to 1 is indicating the aperture shape is circular.

Research Journal of Chemical Sciences E-ISSN 2231-606X
Vol. 6(12), 1-5, December (2016) *Res. J. Chem. Sci.*

Effect of Musi River Pollution on Human Anthropogenic Activities

Srisailam Gogula[1] and Sunder Kumar Kolli[2]*
[1]Department of Chemistry, Govt. City College, Hyderabad, Telangana-500008
[2]Department of Chemistry, Annamacharya Institute of Technology Sciences, Hyderabad
sunderkolli@gmail.com

Available online at: www.isca.in, www.isca.me
Received 22nd June 2016, revised 5th December 2016, accepted 7th December 2016

Abstract

At the present day the world is mainly focused on the depletion of the atmospheric ozone layer by environmental pollution. Environmental pollution is unfavorable alteration of our surroundings. The water is most important resource and one of the universal solvent, it is used by living organisms. The major source of water is mainly oceans, rivers, lakes, ponds and makes 65% of human body. The water is using for daily activities and also used for several industries may causes water pollution. In present study, an extensive investigation of physico-chemical parameters of water samples of river Musi located in Hyderabad was carried out. For this area sampling sites were selected along the river Musi in and around Hyderabad on affected areas like Himayath Sagar-1, Langer House-2, Govt. City college-3, Nagole-4 and Peerjadiguda-5 (Ground water). Water samples were collected during a month of February 2016. The observed values of different parameters such as Colour, Odour, pH, EC, TDS, Turbidity, CO_3, HCO_3, Cl, F, NO_3, SO_4, Na, K, Ca, Mg, TH, BOD and COD of samples were indentified in different locations in and around Hyderabad city.

Keywords: Monitoring, Parameters, Water pollution, Total Dissolved Solids (TDS), BOD and COD.

Introduction

Hyderabad is fifth largest city in India and capital of telangana state with a population of 12 millions. Hyderabad is the creation of the Quthubshahi rulers and located on the Deccan plateau along the Musi river. The physiography of Hyderabad is dominated by hills, monuments, tanks with a rich and varied heritage. Water is natural resource and it gives habitate for diverse types of aquatic life in rivers, lakes and oceans and human body contains two third percentage of water[1]. Produce of waste water is due to increased population, urbanization, domestic, industrial and commercial sectors[2-6]. Day by day the world population increases the demand for food production, industrial activities and domestic purposes grows and leads to heavier withdrawals of the limited renewable fresh water resource[7-9].

The major portion of the wastes disposed off into the atmosphere and the land is washed out by precipitation and runoff, filtration processes and human waste load accumulates in surface water bodies and ground water aquifers. Simultaneously these human activities generate wastes which are discharged into the depleted water resources despoiling them. The industrial wastage as well as domestic sewage/wastes are disposed in the rivers and release of wastes containing wide variety of organic, inorganic pollutants including solvents, Oils, grease, plastics, plasticizers, phenols, heavy metals, pesticides and suspended solids are hazardous substance into rivers, might lead to environmental disturbance[10-12].

The Musi river originated in Ananthagiri Hills located at Vikarabad Ranga Reddy District 90 kilometers to the west city of Hyderabad. The river flows through the Nalgonda district and the water is used for agricultural and horticultural purpose through small reservoirs. In the city of Hyderabad Musi flows through the telangana high court, Govt City College, Osmania general hospital, salarjung museum, state central library, Mahatma Gandhi bus station. Many bridges built on the river, which connects old and new city. The old bridge known as purana pul constructed by the Ibrahim Qutub Shah in 1579AD. New bridge at Nayapul near high court telangana other bridges at Dabirpura, Chaderghat, Amberpet, Nagole and Uppal Kalan in the city of Hyderabad. In the year of 1908 great flood was there on the Musi river. Much damage was there, 80,000 houses damaged and 15,000 people died. That is why Osman Sagar (dam) built by the engineer Nawab Ali Nawaz Jung Bhadur and later on Himayth Sagr dam was builded. The water was used for drinking purpose and through this water one lack hectors is cultivating at the down streams of the river Musi during and after the monsoon rains. In rural area the major crop is rice and grass is the crop in the city. Must River received large scale of untreated sewage from city of Hyderabad through industrial and domestic, disposal dumping sometimes medical wastage also dumped in the river. It is because rapid and uncontrolled urbanization. Due to water demand increased day by day in and around the greater Hyderabad, now drinking water inflows from Krishna, Manjeera and Godavari through the water pipelines to Hyderabad city and waste water release the city is disposed into the Musi river, Due to these reasons water smells unobjectable odour and people afraid to touch the river water. It became 6th

RESEARCH ARTICLE

Ultrasound-Assisted 3-Component Reaction in Acetic Acid Alone: Catalyst / Promoter / Ligand Free Synthesis of Bioactive Pyrazolo[1,5-*a*]pyrimidines

Namburi Suresh[a], Bodapati Veera Durgarao[a], A. Ratnakar[b], Sunder Kumar Kolli[c], Mohd Ashraf Ashfaq[c], Mandava V. Basaveswara Rao[a,*] and Manojit Pal[c,*]

[a]*Department of Chemistry, Krishna University, Machilipatnam, AP, India;* [b]*Department of Chemistry, VR Siddhartha Engg. College, Vijayawada, AP, India;* [c]*Dr. Reddy's Institute of Life Sciences, University of Hyderabad Campus, Gachibowli, Hyderabad 500 046, India*

ARTICLE HISTORY

Received: December 10, 2016
Revised: January 12, 2017
Accepted: January 20, 2017

DOI:
10.2174/1570180814666170126120408

Abstract: Acetic acid alone when employed as a solvent under ultrasound irradiation has been found to be effective for the three-component reaction involving ethyl-5-amino-1*H*-pyrazole-4-carboxylate, aromatic aldehydes and terminal alkynes in the presence of aerial oxygen. The catalyst / promoter / ligand free method afforded a range of pyrazolo[1,5-*a*]pyrimidines as potential and new cytotoxic agents. While some of these compounds showed cytotoxicities against two breast cancer cell lines one of them was found to be promising.

Keywords: Ultrasound, MCR, acetic acid, pyrazolo[1,5-*a*]pyrimidine.

INTRODUCTION

The pyrazolo[1,5-*a*]pyrimidine framework is continuing to be an attractive scaffold for studies in the area of Medicinal Chemistry and drug discovery. Indeed, compounds containing this framework have shown various pharmacological properties including antibacterial, antitumor activities and cytotoxicities [1-3]. Additionally, pyrazolo[1,5-*a*]pyrimidines have shown vast synthetic importance in the preparation of various drugs and bio-active molecules [4-8]. Because of their broad pharmaceutical and synthetic applications, a good number of methods have been developed for the synthesis of compounds containing this *N*-heteroarene [9-17]. For example, 7-(hetero)aryl pyrazolopyrimidines were prepared *via* coupling of 7-chloro-5-phenyl-pyrazolo[1,5-*a*]pyrimidine with (i) arylboronic acids under Suzuki conditions or (ii) an organometallic reagent such as ArZnI in the presence of a Pd-catalyst [18]. This class of compounds has also been prepared *via* $AlCl_3$ mediated C-C bond forming reactions [19, 20].

The single-step cascade reactions such as multi-component reactions (MCRs) [21] have attracted particular interest as these methods often provide an inherently more efficient chemical synthesis of target compounds over the conventional bimolecular reactions. This is exemplified by the development of several MCRs such as Strecker, Passerini, Ugi, Pauson–Khand,Biginelli and Mannich reactions. Indeed, MCRs are considered as powerful tools for the generation of diversity based library of small organic molecules required by both chemical and pharmaceutical R&Ds [22, 23]. Thus study and exploration of MCRs that are practical, effective and scalable have become an important area of research in academic and industrial organizations. Accordingly, the use of MCR in the synthesis of 7-(hetero)aryl pyrazolopyrimidines has been explored by us earlier [24]. While found to be effective this method however required the use of a strong acid catalyst *e.g.* TfOH and temperature 100-110 °C for 2h. Thus the development of a faster and milder method was desirable.

The ultrasound-assisted organic reactions [25-27] have attracted enormous attention due to the efficiency (*e.g.* shorter reaction time, milder conditions, higher yields *etc.*) and greenness (in terms of energy conservation and waste minimization) of these processes over the conventional heating methods. Herein we report a ultrasound-assisted improved synthesis of 5,7-diaryl pyrazolopyrimidines (**4**) in acetic acid alone without using any additional catalyst, promoter or ligand (Scheme **1**). While, the ultrasound-assisted synthesis of pyrazolo[1,5-*a*]pyrimidine having different

*Address correspondence to these authors at the Department of Chemistry, Krishna University, Machilipatnam, AP, India; Tel: +91 40 6657 1500; Fax: +91 40 6657 1581; E-mail: vbrmandava@yahoo.com and Dr. Reddy's Institute of Life Sciences, University of Hyderabad Campus, Gachibowli, Hyderabad 500 046, India; E-mail: manojitpal@rediffmail.com

1570-1808/17 $58.00+.00

Annamacharya institute of technology & sciences

Piglipur, Batasingaram, Hayathnagar (M), Hyderabad, R.R.Dist-501512

3.3.2Number of Research papers per teachers in the journals notified on UGC website during the last five years.

YEAR	2018-19	2017-18	2016-17	2015-16	2014-15
NO.OF.PAPERS PUBLISHED	1	3	0	0	0

Title of paper	Name of the author/s	Department of the teacher	Name of journal	Year of publication	ISSN number
2017-2018					
Experimental Investigations on Performance Evaluation of Four Stroke Copper Coated SI Engine With Methanol Blended Gasoline With Catalytic Converter-A Review	Dr.P.V.Krishana Murthy	MECHANICAL	International Journal of Engineering Science Invention	2017	2319 – 6734
DESIGN AND STRUCTURAL ANALYSIS OF CNG	B.Jeevan kumar	MECHANICAL	International Journal of Advanced Research Trends in Engineering and Technology	2017	2394-3785
Performance And Emission Characteristics of Single Cylinder Diesel Engine with Safflower Biodiesel Blends	P. Sreenivasulu	MECHANICAL	International Journal of Computational Engineering Research	2017	2250 – 3005
A Review on Significant Parameters and Exhaust Emissions of Four Stroke Copper Coated SI Engine with Alcohol Blended Gasoline through Catalytic Converter	Dr. P.V.Krishna Murthy	MECHANICAL	International Journal of Advanced Mechanical Engineering	2018	2250-3234

International Journal of Engineering Science Invention
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 6 Issue 10|| October 2017 || PP. 48-67

Experimental Investigations on Performance Evaluation of Four Stroke Copper Coated SI Engine With Methanol Blended Gasoline With Catalytic Converter-A Review

B.Raja Narender[1] Dr.P.V.Krishana Murthy[2] Dr.Md.Yousuf Ali[3] Dr.M.V.S Murali Krishna[4]

[1]*Associate Professor, Department of Mechanical Engineering,Anurag Group of Institutions (CVSR) Venkatapur, Ghatkesar, Medchal.Dist,Telangana,India*
[2]*Principal, Annamacharya Institute of Technology &, Sciences, Batasingaram, R.R.Dist,Telangana, India*
[3]*Principal, Avanthi Institute of Engineering & Technology,Gunthapally,Hayathnagar,R.R Dist, Telangana,India*
[4]*Professor, Department of Mechanical Engineering,Chaitanya Bharathi Institute of Technology Gandipet, Hyderabad,Telangana, India*
Corresponding Author: B.Raja Narender[1]

Abstract: *The performance evaluation of four stroke spark ignition engine with gasoline and alcohol (methanol and ethanol) blended gasoline was reviewed and research gaps were identified. This paper reports performance evaluation of four–stroke, single–cylinder, water cooled, variable compression ratio (3–9), variable speed (2200–3000 rpm) spark ignition engine with brake power of 2.2 kW at a speed of 3000 rpm. The combustion chamber of the engine was coated with copper (Copper Coated Engine, CCE) [copper-(thickness, 300 μ) was coated on piston crown, inner side of liner and cylinder head]. The engine was fuelled with methanol blended gasoline [20% methanol with 80% gasoline with varied spark ignition timing. The engine was provided with catalytic converter with sponge iron as catalyst along with air injection. The performance of CCE with methanol blended gasoline was compared with engine with conventional combustion chamber (CE) with gasoline operation. Performance parameters of brake thermal efficiency, brake specific energy consumption, exhaust gas temperature and volumetric efficiency were determined at different values of brake mean effective pressure of the engine. Exhaust emissions (carbon mono oxide {CO) emissions, un-burnt hydro carbon (UBHC) emissions and nitrogen oxide (NO_x) levels) were evaluated at full load operation of the engine. Aldehydes (formaldehyde and acetaldehyde) were measured by wet method of 2,4, dinitrophenyle method at full load operation of the engine. Combustion characteristics were measured at full load operation with Piezo electric pressure transducer, TDC (top dead center) encoder, console, and pressure-crank angle software package. NO_x emissions were controlled by employing selective catalytic reduction (SCR) technique with the use of modified zeolite and lanthanum zeolite infused with urea. Methanol blended gasoline operation improved performance, reduced CO, UBHC emissions and NO_x levels when compared with gasoline operation with both versions of the combustion chamber. At recommended and ignition timing, CCE with test fuels of gasoline and methanol blended gasoline improved performance and reduced pollution levels, when compared with CE. Catalytic converter with sponge iron as catalyst along with air injection significantly reduced pollutants with test fuels. Combustion characteristics improved with CCE in comparison with CE with both test fuels.*

Keywords: *Alcohols, copper coating, catalytic converter, fuel performance, exhaust emissions, SCR, combustion characteristics.*

Date of Submission: 04-10-2017 Date of acceptance: 25-10-2017

I. Introduction

In the context of i) fast depletion of fossil fuels, ii) increase of pollution levels with fossil fuels and iii) ever increase of fuel prices in International Market causing economic burden on developing countries like India, the search for alternative fuels has become pertinent. Alcohols (ethanol and methanol) are important substitutes for gasoline as they are renewable in nature and have high octane rating. Methanol can be produced from municipal solid wastes and waste or specifically grown biomass [1]. Though methanol can also be produced from natural gas, there is no point in it as the basic objective is to conserve petroleum gases or liquids. The municipal solid wastes can be converted to methanol. The wastes are first shredded and then passed under a magnet to remove ferrous materials. The iron free wastes are then gasified with oxygen. The product synthesis gas is cleaned by water scrubbing and other means to remove any particulates, entrained oils, H_2 S and CO_2.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Issue 12, December 2017

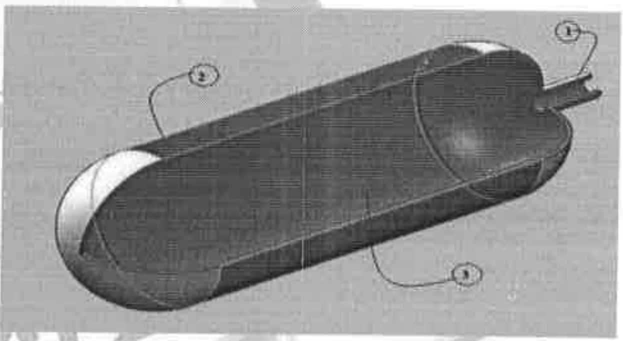
DESIGN AND STRUCTURAL ANALYSIS OF CNG COMPOSITE GAS CYLINDER

B.Jeevan kumar, Asst .Professor, Department of Mechanical Engineering, AITS- Hyderabad
Corresponding author Email-Id:jinnu.20@gmail.com
Dr. M.Madhavi, Professor, Department of Mechanical Engineering, MVSREC- Hyderabad

Abstract: Pressure vessels are essentially storage vessels, but they find a large variety of applications in assorted fields like in industrial processing equipment, where they are subjected to unusual conditions of pressure, temperature and environment. Pressure vessels were constructed from isotropic materials such as steel and aluminum. But now, with the advent of composites In the present study a 70 liters capacity CNG (compressed natural gas) gas cylinder is designed in accordance to ISO 11439:2000(E) standard. The composite gas cylinder is designed for burst pressure 730 bar (73MPa) using netting analysis of filament winding technology. The CNG gas cylinder comprises of cylinder and two end domes, out of one end dome being totally closed. The results indicate the gas cylinder under given loading and boundary condition is safe.

1. INRODUCTION

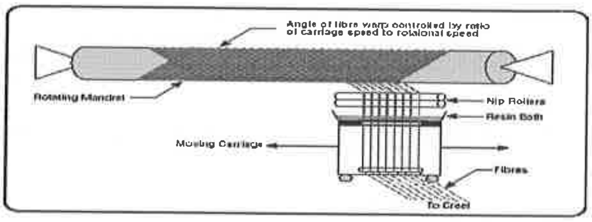
A CNG composite gas cylinder is developed on polymer mandrel using filament winding technology. The aim of the project is to design a 70 liter capacity CNG composite gas cylinder of 340mm diameter and 956mm length in accordance to ISO 11439:2000(E) specifications. Mathematical model is proposed for non-geodesic fiber trajectory on the polymer mandrel. For given geometric specification, ply wise layer design will be done using netting analysis of composite pressure vessel. Further structure analysis is carried out using finite element techniques required for computing failure analysis of composite gas cylinder.

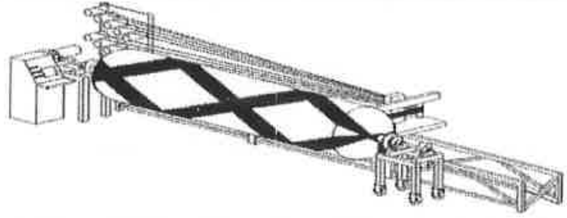


CNG composite Gas Cylinder (1. Metal pole opening, 2. Composite shell ,3. Polymerliner

II FILAMENT WINDING TECHNOLOGY

Filament winding is an automated process in which continuous filament is treated with resin and wound on a mandrel in a pattern designed to give strength in one direction shown in figure filament-wound composite pressure vessels, which utilize a fabrication technique of filament winding form high strength and light weight reinforced plastic parts, are of a major type of high-pressure vessel and are widely used in the


Angle of fibre wrap controlled by ratio of carriage speed to rotational speed
Rotating Mandrel
Nip Rollers
Resin Bath
Moving Carriage
Fibres
To Creel



ISSN (e): 2250 – 3005 || Volume, 07 || Issue, 12 || December – 2017 ||
International Journal of Computational Engineering Research (IJCER)

Performance And Emission Characteristics of Single Cylinder Diesel Engine with Safflower Biodiesel Blends

P. Sreenivasulu[1*],A.V.N.S. Kiran[2]

[1*]Assistant Professor, Department of Mechanical Engineering, Annamacharya Institute of Technology and Sciences, Hyderabad, Ranga Reddy Dist. Telangana State, India.

[2] Assistant Professor, Department of Mechanical Engineering, Annamacharya Institute of Technology and Sciences, Rajampet, Kadapa Dist. A.P, India.

ABSTRACT

The consumption and demand for the petroleum products are increasing every year due to increase in vehicle population, standard of living and urbanization. This causes fast depletion of petroleum products, which leads to move towards the alternative fuels for researchers. Among so many alternative fuels for diesel engine, non edible vegetable oils are the most widely used under the test. The purpose of present work is to investigate the engine performance and emission characteristics of DI diesel engine with different blends of safflower seed oil. Engine tests have been carried out to obtain comparison of fuel consumption, specific fuel consumption, brake thermal efficiency, volumetric efficiency and smoke opacity and compared with that of diesel fuel. The aim of the proposed project work is to investigate experimentally the performance and emissions characteristics of safflower biodiesel at different loads. A single cylinder, direct injection, compression ignition engine has been selected for investigation.

Keywords: D.I.Diesel Engine, Performance, Transesterification, Emissions.

Date of Submission: 14-12-2017 Date of acceptance: 29-12-2017

I. INTRODUCTION

Fossil fuels are one of the major sources of energy in the world today. Their popularity can be accounted to easy usability, availability and cost effectiveness. But the limited reserves of fossil fuels are a great concern owing to fast depletion of the reserves due to increase in worldwide demand. Fossil fuels are the major source of atmospheric pollution in today's world. So efforts are on to find alternative sources for this depleting energy source. Even though new technologies have come up which have made solar, wind or tidal energy sources easily usable but still they are not so popular due to problems in integration with existing technology and processes. So, efforts are being directed towards finding energy sources which are similar to the present day fuels so that they can be used as direct substitutes. Diesel fuel serves as a major source of energy, mainly in the transport sector.

India is importing crude petroleum and petroleum products from Gulf countries. Indian scientists searched for an alternate to diesel fuel to preserve the global environment and to withstand the economic crisis. As far as India is concerned because of its vast agro forestry base, fuels of bio origin can be considered to be ideal alternative renewable fuels to run the internal combustion engines. Vegetable oils from plants both edible and non-edible and methyl esters (Biodiesel) are used as an alternate source for diesel fuel. Biodiesel was found to be the best alternate fuel, technically, environmentally acceptable, economically competitive and easily available. There are more than 350 oil bearing crops that have been identified, among which only sunflower, soyabean, cottonseed, mango seed, rapeseed and peanut oils are considered as potential alternative fuels for diesel engines. Apart from the renewability, the advantages of biofuel are as follows: High oxygen content, higher flash point and higher lubricity that produce complete combustion in comparison with conventional diesel fuel [1]. Further, the environmental benefit is another investigation factor due to lesser greenhouse effect, less air pollution, less contamination of water, soil and reduced health risk [2]. Traditional oilseed feedstock for biodiesel production predominantly includes soyabean, rapeseed/canola, palm, corn, sunflower, cottonseed, peanut and coconut oil [3]. The long chain hydrocarbon structure, vegetable oils have good ignition characteristics, however they cause serious problems such as carbon deposits buildup, poor durability, high density, high viscosity, lower calorific value, more molecular weight and poor combustion. These problems lead to poor thermal efficiency, while using vegetable oil in the engine. These problems can be rectified by different methods that are used to reduce the viscosity of vegetable oils. The methods are transesterification, dilution and cracking method [4]. The transesterification of vegetable oil gives better performance when compared to straight

International Journal of Advanced Mechanical Engineering.
ISSN 2250-3234 Volume 8, Number 1 (2018), pp. 25-38
© Research India Publications
http://www.ripublication.com

A Review on Significant Parameters and Exhaust Emissions of Four Stroke Copper Coated SI Engine with Alcohol Blended Gasoline through Catalytic Converter

[1]D.Baswaraj, [2]P.V.Krishna Murthy, [3]K.Prasanna Lakshmi
[1]Jayaprakash Narayan College of Engineering, Dharmapur, Mahabubnagar. Dist, Telangana
[2]Annamacharya Institute of Technology & Sciences, Batasingaram, Hayathnagar, R.R.Dist, Telangana
[3]JNTU College of Engineering Manthani, Peddapalli Dist, Telangana.

Abstract

In this paper, the performance of four stroke single cylinder spark ignition (SI) engine with copper coated combustion chamber [copper-(thickness,300μm) coated on piston crown, and inner side of cylinder head] with alcohol blended gasoline were investigated. Performance parameters like brake thermal efficiency, exhaust gas temperature and volumetric efficiency at various values of brake mean effective pressure of the engine and also investigated the combustion characteristics such as peak pressure, maximum rate of pressure rise, time of occurrence of peak pressure and maximum heat release at full load operation of the engine with alcohol blended gasoline. In this study, a comprehensive review of the four stroke copper coated spark ignition engine using alcohol blended gasoline with catalytic converter. The output power and emissions of alcohol blended engines were compared with conventional engines with pure gasoline operation.

Keywords: Spark Ignition (S.I) Engine, Conventional Engine (CE), Copper coated combustion chamber (CCCC), Copper coated engine (CCE), Catalytic converter (CC).

I. INTRODUCTION

The performance and pollution levels of four-stroke, single cylinder spark ignition (SI) engine with methanol blended gasoline (20% methanol, 80% gasoline, by volume) having copper coated engine with catalytic converter and compared with conventional SI engine with gasoline operation were studied [1] by M.V.S.Murali

MARKETING STRATEGIES BETWEEN GOODS AND SERVICES – A COMPARATIVE STUDY

[1] P. L. S. Padma Raja Rao [2] Dr.G .Chandra Sekhar

[1] *Research Scholar, Dravidian University*

[2] *Professor & Director, Lords School of Business, Lords Institute of Engineering & Technology, Hyderabad, India*

ABSTRACT

The aim of this study is to compare the marketing strategies between goods and services. Selected companies for goods and services have been considered and their core marketing strategy had been analyzed. The results of this study provide insights for practicing marketing managers while formulating strategic marketing decisions. It is also evident from this study marketing goods is simple whereas marketing services is typical because it requires customer involvement either directly or indirectly.

Keywords: Marketing strategy, strategic decisions, marketing goods, marketing services,durable goods, e-services.

I. INTRODUCTION

Organizations formulate marketing strategies for attaining success in the business. Marketing goods and marketing services are two different things because goods are tangible and services are intangible. Hence each of them requires a unique marketing strategy for reaching the consumers. When an organization develops a new product it is important to remember that goods and services are different and each of them requires unique strategy. The advent of globalization and trade agreements between the nations had given scope for marketing of services and goods across the borders. It would be easily possible to transfer goods to destination and it is happening from centuries but transfer of services from source to destination across the borders is complex process. Only due to support of internet technology some of the services are being marketed at the global level. In addition to services the goods electronic services (e-services) have also evolved in the recent years. A comparison of goods, services and electronic services is shown in Table 1.

A strategy is a long term plan and it is needed to provide direction for all the members in organization during production of goods and services. The marketing strategy is sub plan under the organization level strategy. Based on the product type organizations formulates either goods strategy or services strategy and implement them for effective marketing of products. This paper is an attempt to understand how companies are marketing goods and services and comparing their strategies. The marketing strategy of automobile companies and marketing strategy of selected banks are compared in this regard.

Table 1. Difference between Goods, Services and Electronic Services

Goods	Services	E-Services
• Physical presence	• Cannot be seen	• Cannot be seen
• Patent rights exists	• Not applicable	• Patent can be taken
• Process nature not exits.	• Consists of process nature	• Consists of process nature.
• Can be stored	• Cannot be stored	• Cannot be stored
• Used only with consumption	• Used only with consumption.	• Can be used without consumption.
• Homogenous in nature.	• Heterogeneous	• Homogenous in nature.
• Ownership takes place.	• Non- ownership.	• Non-ownership.
• Dimensions exist.	• No shape	• Measured in bits/bytes

(Source: Compiled by the researcher)

Marketing strategy can be defined as a process of concentrating on firm's capabilities, energies and resources for a course of action which can enhance sales and dominance of a targeted market niche. A marketing strategy integrates product development, promotion, distribution, pricing, relationship management and other elements. A strategic marketing decision considers organization's marketing goals, and explains how they will be achieved, ideally within a stated timeframe.

II. RESEARCH OBJECTIVES

- ➢ To describe about strategic marketing decisions for goods and services.
- ➢ To compare the strategic marketing decisions between goods and services from select companies.
- ➢ To provide some suggestions for practicing marketing managers while developing marketing strategies.

III. SCOPE AND LIMITATIONS

The marketing decisions related to goods and services are compared from the perspective of elements of marketing mix. The companies from India are selected and they belong to automobile industry and banking industry. The companies pricing strategy is given more importance compared to other factors. The time and cost are major limitations else first had information could have been gathered from executives in those organizations.

IV. REVIEW OF LITERATURE

Desai et al (2007) had explained the production of goods and services under uncertain marketing demand. Even though goods can be stored for some time period but they also cannot be produced when there is huge demand because of lead time during the production cycles. The technological advancements also make goods perishable because when innovative product enters into the market the existing products with old technology becomes scrap. For example when Apple Inc had launched iPod then Walk-man and other similar devices have become scrap in the market. Hence storage is an issue for both goods and services when there is uncertain market demand.

According to Ke et al (2012) had stated that marketing of virtual goods and virtual services is rapidly growing with support of virtual infrastructure like e-commerce platforms. The pricing strategies for virtual goods and virtual services are completely different from traditional goods and services. Virtual goods are exchanged between users through online communication and from distant places. Fay and Xie (2008) had described about probabilistic method of selling the goods where distinct items are purchased in multiple sets based on the requirement of consumers.

Taherdoost et al (2014) had explained the marketing of services in international markets through electronic distribution channels. The era of services through internet technology had rapidly grown in the recent years for example patient treatment through video chat. The features of electronic services are process nature, non-rival, self-service and interactive nature in addition to traditional services characteristics. However electronic services can be marketing worldwide whereas traditional services are marketed domestically. In electronic services both service provider and consumer need to have knowledge on usage of technology.

Vargo and Lusch (2008) had stated that organizations are offering services along with goods for attaining the needs of consumers. The firms are shifting from goods dominant position to service and good dominant position for sustaining the competitive business world. It is also observed that most of the economies are exchanging services more than goods in the recent years. Even in business to business market the firms are giving equal importance for services and goods so that they can reach the final customer without intermediaries. Vargo and Lusch (2004) had initially mentioned that organizations developed strategies for marketing goods but later need aroused to developing strategy for marketing services.

Ulaga and Reinartz (2011) had stated that success can be attained by firms if they offer hybrid offerings which mean goods and services should be combined for delighting the customers. Organization should become one stop place for getting all the related services to the product for example automobile companies entering into finance business and insurance business to provide the service to customer at the time of purchase of core product. The capability of manufacturing firms can be enhanced by colleting inputs from the sales force and service related data processing. Organizations posses unique resources like installed product base, manufacturing assets, sales force and field service. The organizations need to develop distinctive capabilities like interpretation capability, design to service and hybrid offering deployment capability.

V. RESEARCH METHODOLOGY

Secondary data from books, journals, magazines, reputed websites and electronic sources had been collected. The information related to selected companies in Electronic industry and Financial services in Service sector had been collected websites of respective companies. The data was compiled in tabular form. The data between the companies in the same industry and between the industries had been compared for this study. The electronic industry had been considered from the perspective of goods and Financial services had been considered from the perspective of services.

Data Analysis

The core marketing strategy of the selected companies had been kept in tabular format and they are analyzed by techniques of content analysis. It is observed that companies in the same industry are giving important one factor among marketing mix elements. The actual marketing strategy of selected companies is very complex but the mission or core objective of the marketing strategy had been considered for this study.

Table 2 Marketing Strategy for Goods of Select Companies

Bajaj	Godrej	LG
To develop products for meeting the requirement of domestic market.	Products for all segments of the market	Product penetration strategy

(Source: Compiled from Secondary Data)

Table 3 Marketing Strategy for Services of Select Companies

Karvy	Broadridge	Angel Broking
Providing Financial Solutions for end to end customers.	Target markets are tech-savvy, entrepreneurs and high net worth individuals.	Targets business class and tech savvy segments of the market.

(Source: Compiled from Secondary Data)

Strategic marketing decisions of selected automobile companies and banks are reviewed for understanding the strategic decisions related to goods and services. From the perspective of good the three automobile companies which are mentioned in Table 2 have developed their strategies based on market segment. They are manufacturing products suitable to various segments of the markets. From Table 3 it is understood that banking firms are more concerned about the needs of individual customers. The service organizations are targeting the customers at the individual level and customizing their product according to their needs. The firms in manufacturing sector are developing products more suitable to the target market segment and not considering the individual needs.

VI. DISCUSSION

It is evident that strategic marketing decisions differ based on product type whether goods or services. In the present era the services are again categorized as general services and electronic services. Goods can be marketed worldwide and does not require customization or involvement of customer. But tough competition exists in marketing of goods and it is also difficult to maintain quality. The brand image plays a vital role in marketing goods therefore allocating high budget to advertisements is essential. For services it is important to maintain efficient front line managers or executives because they represent the brand image of the company.

Customers can have yard sticks to measure quality of products but it is not possible to measure the quality of services. When a quality issue arises with good they can be returned during the guarantee period and later product replacement or refund can be given to customers. But with services if quality is not delivered then it cannot be repaired at the next level. Customer involvement is high in services whereas customer involved may or may not exist in marketing of goods. The profits are more in marketing of services but at the same time there would be high risk. The profits may fluctuate according to the external environment. At the time of recession during 2008 there was low or nil demand for goods but for services the demand is average because services are consumed regularly by consumers.

VII. CONCLUSION

It can be concluded that marketing managers while formulating strategy should give more importance for 'product' in marketing mix which consists of 4Ps. For services marketing managers need to give more importance for 'people' among the 7Ps. The 4Ps and 7Ps list for goods and services respectively is available in Kotler et al (2016). In simple terms it can stated that quality is more important for marketing of goods whereas customer satisfaction or customer involvement is more important for marketing of services.

VIII. FUTURE RESEARCH

In this research the comparison is done between Consumer industry and Financial services in service industry with regard to marketing strategy. The future researchers can compare other industries to examine the strategic decisions between goods and services. The companies belonging to goods and services belonging to Indian origin have been considered future researchers can study the strategies of international companies marketing strategies. There are global companies like GE which is into both manufacturing of goods and services. The strategic marketing decisions of such companies should be compared between goods and services.

REFERENCES

1. Desai, P. S., Koenigsberg, O., & Purohit, D. (2007). Research note—the role of production lead time and demand uncertainty in marketing durable goods. Management Science, 53(1), 150-158.
2. Fay, S., & Xie, J. (2008). Probabilistic goods: A creative way of selling products and services. Marketing Science, 27(4), 674-690.
3. Ke, D., Ba, S., Stallaert, J., & Zhang, Z. (2012). An empirical analysis of virtual goods permission rights and pricing strategies. Decision Sciences, 43(6), 1039-1061.
4. Kotler, P., Keller, K. L., Brady, M., Goodman, M., & Hansen, T. (2016). Marketing management. Pearson Education Ltd..
5. Taherdoost, H., Sahibuddin, S., & Jalaliyoon, N. (2014). Features' evaluation of goods, services and E-services; electronic service characteristics exploration. Procedia Technology, 12, 204-211.
6. Vargo, S. L., & Lusch, R. F. (2004). The four service marketing myths: remnants of a goods-based, manufacturing model. Journal of service research, 6(4), 324-335.
7. Vargo, S. L., & Lusch, R. F. (2008). From goods to service (s): Divergences and convergences of logics. Industrial marketing management, 37(3), 254-259.
8. Ulaga, W., & Reinartz, W. J. (2011). Hybrid offerings: how manufacturing firms combine goods and services successfully. Journal of marketing, 75(6), 5-23.