IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 11, Issue 4 (May. - Jun. 2013), PP 01-07
www.iosrjournals.org

Simulation of Route Optimization with load balancing Using AntNet System

Ashish Kumar Sharma[1*], Baijnath Kaushik[2], A.K. Kohli[3], Navdeep Kaur[4]
[1]*Department of Computer Science, Krishna Engineering College, Ghaziabad, India*
[2]*Department of Computer Science, Krishna Engineering College, Ghaziabad, India*
[3]*Department of Electronics & Communication, Thapar University, Patiala, India*
[4]*Department of Information Technology, Chandigarh Engineering College, Chandigarh, India*

Abstract: *This paper is based on analysis of the performance of load balancing and route optimization in computerized networks. The complete system model shows the scenario of Packet distribution between nodes, and if congestion occurs due to traffic then Packet to be failed. The model used the AntNet system for simulate the network. The simulation runs on the ant's behavior for the load balancing of the network. The ants travel across the network between alternative chosen pairs of nodes; as they travel they deposit pheromones from their source node, collect the information of the route and the traffic congestion encountered on their journey. They select their path at each next node according the distribution of pheromones at each node. Packets between nodes are routed of the pheromone distributions at each next node. The performance of the network is proportional to packets which are failed. This model also shows the adaptivity of the system; the nodes are removed from the network, system finds the alternative chosen paths without system degradation and controls the performance of routing.*

Keywords- *AntNet Algorithm, Ant Colony Optimization (ACO), Routing, Load Balancing, Dynamic, Adaptive, Simulation, Communication Networks.*

I. Introduction

In the communication network; congestion, delay and reliable communication play an important role to improve the performance and optimize the network and the network should be robust and adaptive. Optimization and performance are major issues in the industrial world as well as in the science world. So, these Issues occur in the communication network. Routing is the basic term of network because it handles the information as well as the control system performance. Routing is an important aspect of communication networks because it can greatly influence the overall network performance [3]; good routing can cause greater throughput or lower average delays and all other conditions would be same.

Load balancing [7] is essentially the construction of call-routing schemes which successfully distribute the changing load over the system and minimize lost calls. Of course it is possible to determine the shortest routes from every node to every other node of the network. In this way the average utilization of nodes will be minimized, but this is not necessarily the ideal way to avoid node congestion, idea should has to do with how the traffic on the network is distributed.

Routing is a distributed approach which is responsible for the routing tables, made on each node in the network, which give the information about incoming packets and which next link is used to continue the communication towards the destination node. Routing evolved some issues like user requirement, confliction between nodes and impact of different constraints of different technologies.

The AntNet is based on the ACO [14], which is a routing protocol for packed switched network invented by M. Dorigo and G. Di Caro [1]. ACO elaborates the nature and behavior of ants in a colony and could find the shortest path as well from the trail of the pheromone (chemical substance) deposited by the other ants. Through pheromone, ants could find out the shortest path using shortest path mechanism. After efforts of researchers on ants, a routing algorithm generated; which was realistic, and comparatively better than other algorithm. ACO [6, 7, 8] improves the performance of the network and as well optimize with load balancing of the network. The problem of the network represented by the node graph, where artificial ants construct multiple paths from the source node to destination node and trace the path until optimal solution not found between two end node and it could take number of iteration in tracing of paths. Artificial ants collect the quantitative and qualitative information path cost and load of traffic in the network .These fundamentals are based on the ant colony system (ACS). ACS concept is used in AntNet [2]. An efficient routing algorithm will minimize the number of nodes that a packet will need to connect to in order to be completed thus; minimizing network load and increasing reliability. An implementation of AntNet based on Marco Dorigo and Thomas Stutzle has been designed.

II. Issues In Communication Network

Routing is basic concept in communication network. It works to address the traffic between source node to destination node with determines the network performance. If routing has such importance then routing algorithms have the goals of directing data traffic from source to destination nodes in measure of network performance. The basic problem of routing is, to find the path of minimum cost between any two nodes that could be source node to destination node. Routing algorithms can be divided into three different categories; non-adaptive, adaptive, distributed. Non-adaptive routing algorithm is static routing, offline using local information, routing is not according to the traffic of network and the network topology. Adaptive routing [3, 4, 5] algorithms are dynamic and centralized routing, to produce the optimal solution from the collected information through the entire network using global information. Distributed routing algorithms have mixture of both above algorithms that mean, uses mixture of offline local and global information. Routing algorithms strongly interact with congestion and admission control algorithms to determine the overall network performance.

The network performance depends on few constraints as like throughput (data delivered in unit time) and packet delay (sec). Throughput quantifies the quantity of service that the network has been able to offer in a certain amount of time, while packet delay defines the quality of service produced. The precise balance between throughput and average delay will be determined by the flow control algorithms operating concurrently with the routing algorithms. Along with these constraints; less congestion, high delivery rate, produce unexpected great performance, less overheads and less failure are also important measurements in the account.

In the network, problem associated define in the routing is as follows:
1. Find out the shortest path in the routing.
2. Selected shortest path might not be a minimum cost route.
3. Increase the through-put performance with less overhead.
4. Lacks of the least hop count.
5. Less effective congestion control, data flow is complex.
6. Load Balancing, Adaptive network

One of the issues with network routing (especially in very large networks such as the WAN, INTRANET etc.) is adaptability. Not only traffic can be unpredictably high, but the structure of a network can change as old nodes are removed and new nodes added. This perhaps, makes it almost impossible to find a combination of constant parameters to route a network optimally.

In this paper, we are concerned about the load balancing of the communication network with less failure of network and optimize the flow of traffic around the network using routing algorithm. According to the problems of network, optimize the network using AntNet [9, 10] and compare it a generic algorithm which is showing that AntNet performs better than another routing algorithm.

III. AntNet: Adaptive Routing Algorithm

The AntNet was proposed by M. Dorigo in 1997 and with modification in 1998 [1, 2, 5]. This paper is based on working of AntNet and solution implemented on it.

The operation of AntNet is based on two types of agents:

- Forward Ants who grab information about the status of network
- Backward Ants who grasp the information manipulate and update the routing tables of routers which is already situated on network.

The AntNet algorithm describes in steps:

1. On a regular time base, each source node sends a Forward Ant to a random destination.
2. In each node they pass, the elapsed time since the start is stored on an internal stack together with the identifier of the node. The Forward Ant selects the next hop in two way
   - It draws between random nodes, where each random node has a probability to be selected.
   - If the node is selected in previous way was already visited then it draws again the jumping node but now with the same probability for all neighbors.
3. If the selected node was already visited then a cycle is found. The Forward Ant pops from its stack all data of the cycle nodes, because the optimal path must not have any cycle.
4. When the Forward Ant reaches its final destination again the elapsed time since the start and the identifier of the node are stored on the stack of the ant. The Forward ant is converts into Backward Ant and transfers its stack. Backward Ant will return to source node with same path in opposite direction.

5. The Backward Ants Use the information collected by the Forward Ants to update the random node routing table along the path. According to data carried in the stack increasing probabilities associated to the path used and decreases other paths probabilities.
6. When updation has been done and Backward Ant arrives on source node, stop working.

Due to algorithm, shows that Backward Ants have a higher priority than Forward Ants because Backward Ants has to have processes as fast as possible to make algorithm more adaptive comparison Forward Ants suffers from delays so that algorithm can fight with Congestion.

For Load balancing, the times that are saved onto the stack of the Forward Ants, are not computed as the difference between two timestamps. They are computed as the sum of two terms, one representing the delay due to link load, and one for the node load. When the load on a link becomes very high; the term representing the link load in increased to improve the load balancing [11] over multiple paths.

IV. Expermental Settings

The bi-directional, un-weighted topological network of 30-node was chosen because this is a realistic interconnection structure of a possible switch-based network. It is also the same topology as was used in (Appleby & Steward, 1994) [12], and is in fact the structure of the British Synchronous Digital Hierarchy (SDH) network.

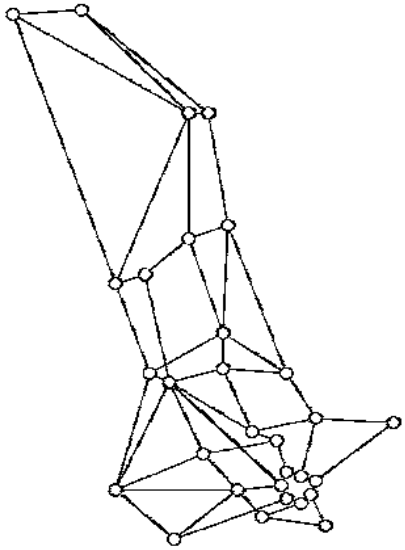


FIG- 1 The network topology is same as the SDH network of British Telecom and used in simulation.

The network is most naturally represented by an undirected graph. Each node in the graph corresponds to a switching station; the links between nodes correspond to communication channels. A given node will usually only be linked to a subset of other nodes, usually its geographical neighbors; links are bidirectional.

Every time step of the simulation proceeds as follows. First, packets that have expired are removed, releasing capacity at the nodes. Next, packets are sent by a traffic generator. These packets included a source node, a destination node and duration, measured in time steps. When a packet is sent, its route is determined by the current routing tables. The source and destination nodes are randomly chosen as a function of each node's probability of being an end node; this is both convenient and reasonable.

The simulator projected in Microsoft visual studio 2008 with the help of SQL server 2005 for the database of report. The complete simulation is based on the AntNet algorithm and following the criterion according the algorithm, whether it is related with parameters, design of network or working. In fact, simulator gives facility to examine the simulation through report of both AntNet On and Off mode and graph representation is plus point to compare the both result of modes. From running simulator, given tactics of simulation with respect of on and Off mode of AntNet [13]. The simulation having different tactics like speed (ticks/sec is a unit of near speed as system can run), packets, node capacity, concurrent packets and packet duration etc. Calculation will work on these fields of simulation.

V. Results

The experimental results are showing in the tables 1to 2 and graphs 2 to 4; which are illustrate how the AntNet algorithm affects the routing of traffic. These tables showing the effectiveness of the algorithm against the system running without AntNet. Since it is possible to switch nodes on and off, a number of test

comparisons will be done to show how AntNet can improve the routing of a network when paths are no longer valid and new routes have to be chosen.
When AntNet Simulator switched off that mean none of the pheromone values are read or written and work in generic algorithm.

The simulation ran numbers of times on few parameters. The Average best results obtained when values were changed with each and every simulation. The parameters values were as follows:-

i. Total packets- 10, 50, 100, 150, 200
ii. Concurrent Packets- 4, 20, 40, 80, 120
iii. Node Capacity- 6, 30, 60, 100, 150
iv. Packet duration- 999 (same)

These parameters can run on different simulation speed as 1, 5, 10, 50, 100, 1000 ticks/sec which is almost similar to system speed.

Results shows with AntNet On and AntNet Off as follows:

- AntNet On gave better results than AntNet Off in Throughput (total packets send).
- AntNet On gave better results than AntNet Off in Current average packet.
- AntNet On gave better results than AntNet Off in Distance.
- AntNet On gave better results than AntNet Off in Delay.

	AntNet ON	AntNet Off
Current Average Packet	6.49	6.08
Total Packets Send	100	102
Distance	185.69	267.35
Delay	.18569	.26735
Failed Packets	0	2

Table 1 When Simulation Speed =1000 and Total Packets =100

Results are depends upon the simulation speed which is mention in above table and correspondingly mention different results which is elaborated that AntNet On is better than AntNet Off. Different results values is showing that whether , it's about total packet send over network, Distance and delay and failed packet but giving average best result over AntNet Off.

	AntNet ON	AntNet Off
Current Average Packet	6.11	5.86
Total Packets Send	100	101
Distance	154.28	186.95
Delay	1.5428	1.8695
Failed Packets	0	1

Table 2 When Simulation Speed =100 and Total Packets =100

Results is showing in table based on when simulation speed is 100 and total packets are same in previous table and still AntNet On providing average best result than AntNet Off.

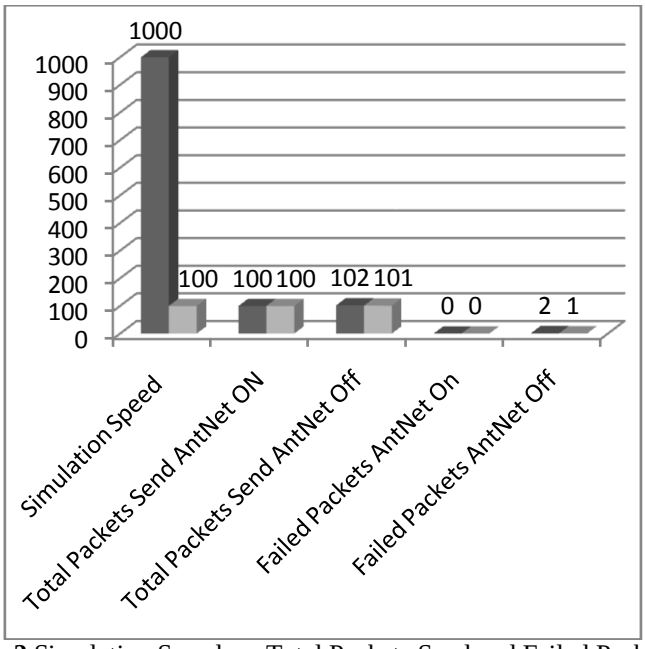


Fig.2 Simulation Speed vs. Total Packets Send and Failed Packets

In the Fig. 2, various results shown according simulation speed and differentiate how many packets send and how many packets failed.

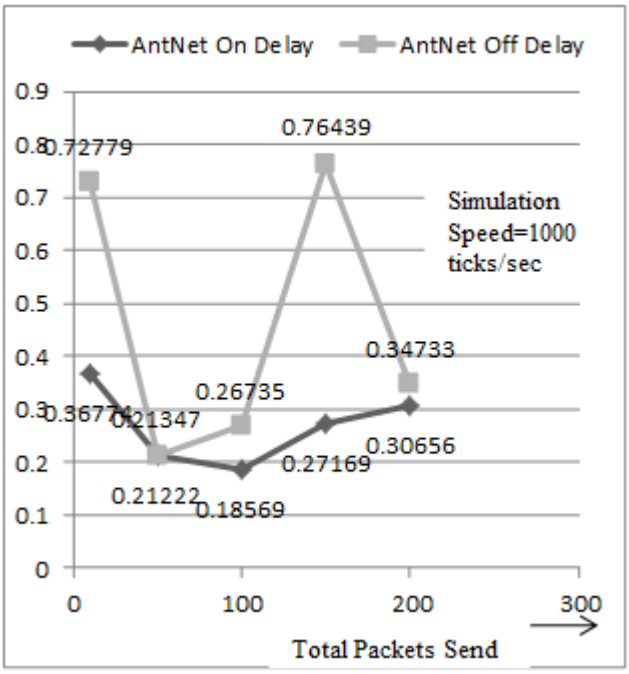


Fig. 3 Comparisons between Delay of AntNet On and Off.

In the Fig. 3, Packet delay is shown of both mode when simulation speed is 1000. Packet delays on nodes are less in AntNet On comparison in AntNet Off.

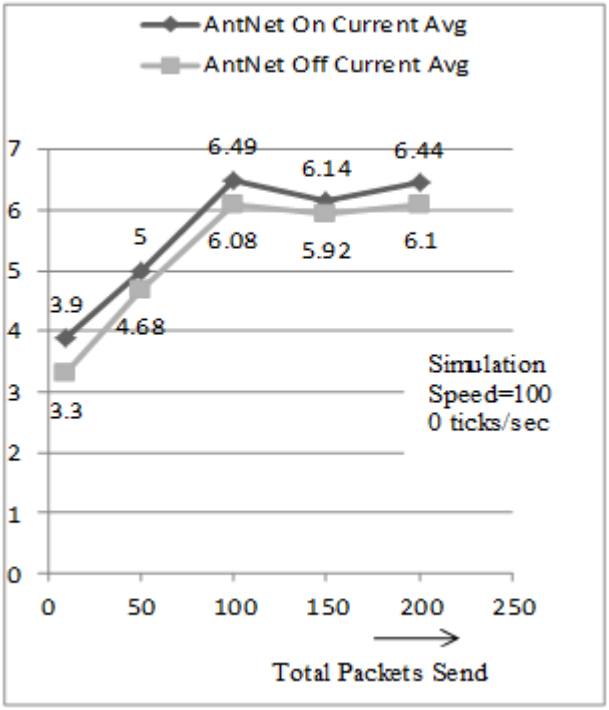


Fig.4 Comparisons between Current Average of AntNet On and Off.

In the fig. 4, Current Average packet sending is explaining that when we use AntNet On then its provides better results than AntNet Off.

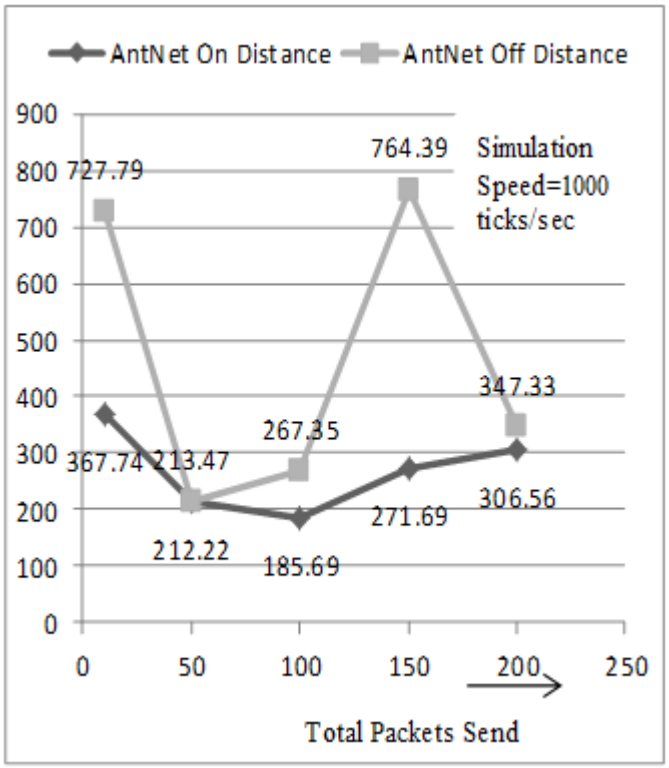


Fig. 5 Comparisons between Distance of AntNet On and Off.

In the Fig. 5, is same exactly like delay comparisons but here mention the distance between two nodes which is less in AntNet mode On.

VI. Conclusion And Future Works

The Result clearly shows that AntNet algorithm performs much better than Non AntNet algorithm and ant based load balancing fulfill technique .This was the basically a concept of packet switched telephony used in SDH British Telecom network. Further work could be that this simulation will work with real packet switched telecom system or VoIP. This simulation can be modified for more large number of nodes means it can be enhanced in future. In this scenario many concept remaining like evaporation, size of packets, and reaction of simulation after packet size etc. If we execute the simulation for a long time that means over a period of hours then processing of system reflects in a long way this concept is evaporation is remaining in and it can be resolute in future.

References

[1] G. Di Caro, M. Dorigo: *AntNet: A Mobile Agents Approach to Adaptive Routing. Technical Report, IRIDIA, 1997.*

[2] G. Di Caro, M. Dorigo: *An adaptive multi-agent routing algorithm inspired by ants behavior. IRIDIA, Belgium, 1998.*

[3] Benjamin Baran, Ruben Sosa: *AntNet routing algorithm for data networks based on Mobile Agents. Inteligencia Artificial, 2001.*

[4] G. Di Caro: *Ant Colony Optimization and its Application to Adaptive Routing in Telecommunication Networks, IRIDIA, 2004*

[5] G. Di Caro, M. Dorigo: *AntNet: Distributed Stigmergetic Control for Communication Networks, Journal of Artificial Intelligence Research, 9, Vol. 9, PP. 317-365, IRIDIA, 1998.*

[6] Benjamin Baran, Ruben Sosa*: A New approach for AntNet Routing, Presented at the Proc. 9th Int. Conf. Computer Communications Networks, Las Vegas, NV, 2000.*

[7] Ruud Schoonderwoerd, Owen Holland, Janet Bruten, Leon Rothkrantz: *Ant-based load balancing in telecommunications networks, HP lab Bristol, U.K., Tech. Rep. HPL, 1997.*

[8] E. Bonabeau, Florian Henaux, Sylvain Guerin, Dominique snyers, Pascale Kuntz, Guy Theraulaz: Routing in telecommunication networks with smart ant-like agents, Santa Fe Institute, 1999.

[9] V. Verstraete, M. Strobbe, E. Breusegem, Jan Coppens, Mario, Pickavet, and Piet Demeester: *AntNet: ACO routing algorithm in practice, Ghent University, Belgium, 2002.*

[10] M. Dorigo, G. Di Caro and L. M. Gambardella: *Ant Algorithms for Discrete Optimization. Artificial Life, MIT Press, 1999*

[11] S. Lipperts, Birgit Kreller: Mobile Agents in Telecommunications Networks – A Simulative Approach to Load Balancing, German Research (DFG), 1999.

[12] S. Appleby, S. Stewards: *Mobile Software agents for control in Telecommunication Networks, BT Technology Journal, Vol.12, no.2, 1994.*

[13] Suchita Upadhyaya, Richa Setiya: *Ant Colony Optimization: A modified Version, ICSRS, 2009.*

[14] M. Dorigo, M. Birattari, T. Stutzle: *Ant Colony Optimization IRIDIA, Technical Report, TR/IRIDIA/2006-023, September 2006.*

IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 11, Issue 4 (May. - Jun. 2013), PP 08-12
www.iosrjournals.org

Deniable Encryption Key

P.Lokesh Kumar Reddy1[1], B.Rama Bhupal Reddy[2], S.Rama Krishna[3]
[1]*Rama Raja Institute of Technology and Science, Tirupati, A.P., India*
[2]*Dept. of Mathematics, K.S.R.M. College of Engineering, Kadapa, A.P., India*
[3]*Dept. of Computer Science, S.V. University, Tirupati, A.P., India*

Abstract: *Deniable encryption is an important that allows a user (a sender and/or a receiver) to escape a coercion attempted by a coercive adversary. Such an adversary approaches the coerced user after transmission forcing him to reveal all his random inputs used during encryption or decryption. Since traditional encryption schemes commit the user to his random inputs, the user is forced to reveal the true values of all his random inputs (including the encrypted/decrypted messages and the encryption/decryption keys) which are verifiable by this coercer using the intercepted cipher text. In this scenario, a coercer may force the user to perform actions against his wish. An appealing property in the mediated RSA, PKI was introduced that, the user has no information, neither about his full private (decryption) key, nor the factorization of the RSA public modulus, which represents an excellent step toward achieving in forcibility in public key encryption, since, a coercer cannot ask the user to reveal such unknown information. In this pa- per we present a scheme for receiver-deniable public-key encryption, by which, the receiver is able to lie about the decrypted message to a coercer and hence, escape a coercion. On one hand, the receiver is able to decrypt for the correct message, on the other hand, all the information held by the receiver, when opened to a coercer, do not allow this coercer to verify the encrypted message and consequently, approaching this user becomes useless from the very beginning.*

Keywords*: Deniable encryption, mediated PKI, oblivious transfer, public-key encryption, RSA*

I. Introduction

While traditional encryption schemes protect the privacy of the sender and the receiver against eavesdroppers (semantic security), they fail to provide protection against coercers. A coercive adversary has the power to approach the user (the sender and/or the receiver) after the ciphertext is transmitted and of course recorded by this adversary. It commands the user to reveal all his random inputs used during encryption or decryption. Since the cipher text produced, using standard encryption schemes (specially, public-key encryption) commits the user to his random inputs, It cannot lie about the true plaintext. Such commitments allow the coercive adversary to verify the validity of the opened message. Deniable encryption allows a user to escape coercion. Namely, if this user opens all his inputs (including the claimed encrypted message) to a coercer, the coercer fails to prove the validity or invalidity of the opened message.

Deniable encryption maybe classified according to which party is coerced: sender-deniable encryption schemes are resilient against coercing the sender. The Definitions for receiver-deniable and sender-receiver- deniable follow analogously. When the sender and the receiver initially share a common secret key, this is spoken off as shared-key deniable encryption. In deniable public- key encryption, no pre-shared information and no communications are assumed prior to the encryption process. This follows from the assumptions of standard public-key encryption schemes. Yet, deniable public-key encryption is more challenging than deniable shared-key encryption since the public key of the receiver is already known to everyone including the coercer, consequently, neither the sender nor the receiver can lie about the receiver's public key.

The work in [5] showed that it is possible by simple tricks to transform any sender-deniable encryption scheme to a receiver-deniable encryption scheme and vice- versa. Also, they showed that, with the help of other parties with at least one of them remains un-attacked, it is possible to transform a sender-deniable encryption scheme to a sender-receiver-deniable encryption scheme.

In our recent work of, we devised a sender-deniable public-key encryption based on quadratic residuosity of a composite modulus and showed how to device a sender- deniable public-key encryption from any trapdoor per mutation. However, when the schemes are transformed to be receiver-deniable using the tricks of [5], the schemes are no more one-move schemes. Considering only one- move schemes, receiver deniability is more challenging than sender-deniability since in the later case, everyone knows the public-key of the receiver but the private key of the receiver is known only to the receiver who is beyond the reach of the coercer. In the former case, the receiver maybe coerced to reveal his private key which is verifiable using the public key and a dummy message.

Deniable encryption is very useful in the protocols where coercive adversaries come to play as a potential threat. For example, deniable encryption protects voters from being coerced during electronic elections [6, 9]. It is also very useful to protect bidders in electronic auctions. Generally, deniable encryption is very important when a party is forced to act against his/her wish.

Our construction assumes the existence of the simple and efficient mediated-RSA (mRSA) [2] as the PKI in place. mRSA was invented as a method to achieve fast revocation in RSA PKI. mRSA involves a special entity, called the SEM (SEcurity Mediator), an on-line partially trusted server, to help signing or decrypting messages. The CA generates the private key d corresponding to Bob's (the receiver's) public key e and splits this private key into two pieces. One piece (d_{SEM}) is delivered to the SEM and the other piece (d_{Bob}) is delivered to Bob. The pair (e, N) is the usual RSA public key. An important property to notice here is that Bob himself has no information neither about his full private key, nor about the factorization of the public modulus N . This property is an excellent step toward achieving deniability since, a coercer will not ask Bob to reveal such unknown information. To decrypt a received cipher text, C, each party (Bob and SEM) performs his/her partial decryption on C; finally the partial decryptions are combined to recover the plaintext message M. To revoke Bob ability to sign or de- crypt messages, the CA instructs the SEM to stop issuing partial decryptions or signatures (spoken of as tokens) for Bob public key. At this instant, Bob's signature and/or decryption capabilities are revoked. The functionality is equivalent to (and indistinguishable from) standard RSA due to the fact that the splitting of the private key is transparent to the outside, i.e., to those who use the corresponding public key. Also, knowledge of a half-key cannot be used to derive the entire private key. Therefore, neither Bob nor the SEM can decrypt or sign a message without mutual consent.

As our PKI is established, we turn to discuss our tools. To complete the deniability service, we need an efficient protocol for (1-out-of-n) oblivious transfer (OT_1). Rabin proposed the concept of oblivious transfer (OT) in the cryptographic scenario. In this case the sender has only one secret bit b and would like to have the receiver to get it with probability ½, on the other hand, the receiver does not want the sender to know whether it gets b or not. For OT_1 , the sender has two secrets b_1 and b_2 , the receiver will get one of them at the receiver's choice.

The receiver does not want the sender to know which bit he chooses and the receiver must not know any information other than he chosen. Oblivious transfer is a fundamental primitive in many cryptographic applications and secure distributed computations and has many applications such as private information retrieval (PIR), fair electronic contract signing, oblivious secure computation, etc. Our proposed receiver-deniable public-key encryption scheme requires one invocation of OT_1 between Bob and the SEM. Also we need a Secured mechanism [10] to make the OT protocol deniable.

The rest of the paper is organized as follows: Section 2 describes the related work. Section 3 gives our motivations and contributions. The underlying PKI and the oblivious transfer protocol are described in Section 4.

Section 5 states our assumptions and model. . The strong version is given in Section 6. Section 7 shows the techniques to transform deniability. The bandwidth is improved in Section 8. Finally, the conclusions are given in Section 9.

II. Related Work

The work in [5] constructed a sender-deniable public- key encryption scheme based on trapdoor permutations. However, the scheme falls short of achieving an appropriate level of deniability, that is, to achieve a high deniability, the size of the cipher text corresponding to a one bit encryption is super-polynomial and hence inefficient. In the deniable public-key encryption scheme of [5], a one bit plaintext requires tn bits of cipher text where t is the bit-length of elements in a translucent set St and $t = s + k$ for security parameters n, s and k.

The scheme provides deniability of $4/n$ *and* decryption error of $n2^{-k}$. Hence, to achieve a high level of deniability and a sufficiently low decryption error, the cipher text is super- polynomial and almost impractical [5]. Constructed two deniable public-key encryption schemes based on translucent sets, the first represents the building block for the second which they have called, the "Parity Scheme". The work in [5] also notified that in order to build a one-round scheme, different approaches are required. Also, [5] introduced techniques for the less challenging, deniable shared- key encryption and showed that the one-time-pad is a perfect deniable shared-key encryption.

Based on the sender-deniable public-key encryption, the work in [4] described a general multiparty computations allowing a set of players to compute a common function of their inputs with the ability to escape a coercion. In fact, deniable encryption has an impact on designing adaptively secure multiparty computations [4] since, the notion of deniability is stronger than the notion of non-committing encryption Schemes.

III. Motivations and Contributions

In this Section we describe our motivations and contributions of the work in this paper.

3.1 Motivations

Deniable public-key encryption is a strong primitive, essential in all cryptographic protocols where a coercive adversary comes to play with high potential. Deniable public-key encryption realizes the "Receipt-freeness" attribute which is a very important attribute in electronic voting, electronic bidding and auctions. The schemes proposed in [5] fall short of achieving the desired level of deniability and correctness unless the size of the cipher text corresponding to a one bit encryption is super- polynomial. An appealing property in the mRSA PKI [2] is that the user himself has no information neither about his full private key, nor about the factorization of the public modulus N, consequently, a coercer will not ask the user for such unknown information.

3.2 Contributions

The contributions of this paper are to introduce an efficient receiver-deniable public-key encryption (RD-PKE) scheme. Our proposed scheme enjoys the following properties:

- It is a one-move scheme without any pre-encryption information required to be sent between the sender and the receiver prior to encryption.
- No pre-shared secret information is required between the sender and the receiver.
- Achieves a high level of deniability equivalent to the factorization of a large two-prime modulus.
- No deciphering errors.
- The bandwidth (cipher text bit-length) is significantly improved compared to previous constructions.

Efficiency, We reduce the required bandwidth (cipher- text bit-length) to, 2 lg N bits for a single bit encryption, where N is a two-prime RSA modulus. Moreover, this bandwidth can be efficiently improved, that is, 2 lg N bits of cipher text allow about lg $N - \delta$ bits of plaintext encryption where δ is a short randomizing string. At the same time, our scheme provides strong deniability (i.e. undetectable cheating) equivalent to the infeasibility to factor a sufficiently large two-prime modulus. Unlike the schemes of [5], our scheme produces no decryption errors and hence, more reliable. We introduce two versions of our RD-PKE scheme, a weak version to declare our idea and security proofs, and then we show a simple modification to improve this weak version to be a strong RD-PKE scheme.

IV. Preliminaries

4.1 Mediated RSA

Mediated RSA was invented as a simple method to achieve fast revocation in public-key cryptosystem. As usual, a trusted certificate authority (CA) sets up the RSA modulus N, the public exponent e and the private exponent d for the user. Next, instead of delivering d to the user, the CA splits d into two pieces d_{SEM} and d_{user} such that $d = d_{SEM} + d_{user} \bmod \phi(N)$ where $\phi(N)$ is the RSA Euler totient. Finally, the CA secretly delivers d_{user} *to* the user and d_{SEM} to the SEM.

Encryption, For Alice to encrypt a message M $\in Z_N$ to Bob, she uses Bob's public pair (N, e) to compute the usual RSA cipher text $C = M^e \bmod N$ and sends C to Bob.

Decryption, on the reception of C by Bob, the decryption process is as follows:

- Bob delivers C to the SEM.
- If Bob's key is revoked, the SEM returns ERROR and aborts, else,
- The SEM computes her partial decryption $P\ D_{SEM}$ = C $d_{SEM \bmod N}$ and returns P D_{SEM} to Bob.• Bob computes his partial decryption $P\ D_{Bob} = C\ d_{Bob}$

Mod N *and* extracts $M = P\ D_{SEM}\ P\ D_{Bob \bmod N}$.

It is important to notice that the SEM gains no information about the decrypted message M *[*2].

4.2. Oblivious Transfer

Our proposed RD-PKE requires that Bob involves with the SEM in an OT1 invocation to get his encrypted bit. The main objective of the oblivious transfer protocols was to improve the efficiency and security of the protocols. The protocols of [18] have several appealing properties.

First, they prove efficiency over previous protocols, second, there are no number theoretic constrains on the strings to be obliviously transferred, third, the protocols have bandwidth-computation tradeoffs which make them suitable for variety of applications. The protocols of operate over a group Z_q of prime order, more precisely, G_q is a subgroup of order q of Z_* where p is prime and $q/p - 1$. Let g be a generator group and assume that the Diffie-Hellman assumption holds. In their OT1: The sender owns two strings r_0 and r_1. He chooses a random element $U \in Z_q$ and publishes it. The chooser picks a random $1 \leq k \leq q$ and sets $pk\sigma = gk$ where$\in \{0, 1\}$ is the chooser's choice.

The chooser also computes pk1$-\sigma = U/pk\sigma$ and sends pk0 to the sender. The sender picks a random R and computes gR *and U R*, he also computes pk^R and $pk^R = U^R/pk^R$. The sender computes $pk_1-\sigma = U/pk\sigma$ and sends pk_0 to the sender. The sender picks a random R and computes gR *and U R*, he also computes pk_R and $pk_R = U^R/pk^R$. The sender The chooser selects a random k and sets $pk\sigma = gk$ where $\sigma \in \{0, ..., n-1\}$ is his choice, it holds that $pk^i = U^i/pk^0 \ \forall i = (1, ..., n-1)$. The chooser sends pk_0 to the sender. The sender computes pkR *as* well as pk^R

V. Assumptions and Model

We define a *receiver-deniable public-key encryption (RD- PKE) scheme* as a scheme by which, the receiver is able to lie about the decrypted message to a coercer and hence, escape coercion. On one hand, the receiver is able to decrypt for the correct message, on the other hand, all the information held (or extractable) by the receiver when opened to a coercer, do not allow this coercer to verify the encrypted message and consequently, approaching the receiver becomes useless from the very beginning.

The participants in our scheme are the certificate authority (CA), the security mediator (SEM), the sender (Alice), the receiver (Bob) and the coercive adversary (coercer). As usual, the CA is assumed to be fully trusted by all participants. The SEM is a semi-trusted party in the sense that it follows the execution steps word for word but it is willing to learn any information that could be leaked during execution. Alice is assumed to be beyond the reach of any coercer while Bob is possibly coerced.

The coercer has the power to approach Bob coercing him to reveal the decrypted message, the decryption partial key and all the parameters he used during decryption.

This paper describes the scheme allowing one bit encryption at a time. The reader will notice that the scheme can be easily adapted to allow multiple bits encryption at a time. We assume that an mRSA PKI is already in place. Hence, the pair (e, N) represents Bob's public key while d_{Bob} (respectively d_{SEM}) are the pieces of Bob's private key d held by Bob (respectively the SEM). Let *bt be* the true bit to be encrypted by Alice to Bob. The scheme is described next.

Encryption. To encrypt the bit *bt* to Bob, Alice proceeds as follows:

• Picks a lg *N bits* string R $\in_R ZN$. Let $r_0 ... r_{n-1}$ be the binary representation of *R*.
• Scans the binary representation of *R for* an index (Pointer) *i* such that $r_i = b_t$
• Computes and sends the two encryptions, $C_i = i \mod N$ *and* $C_R = R^e \mod N$ to Bob.

VI. Full Deniability

In this Section we show how to achieve full deniability in our RD-PKE scheme, i.e., the scheme will be deniable even if the coercer is capable of eavesdropping the Alice-Bob channel and the SEM-Bob channel as well. The problem is that the OT protocol is not deniable and hence commits Bob to what he receives from the SEM. We benefit from the fact that the SEM and all its users are in the same domain (or System) this fact facilitates the sharing of a time-synchronous pseudo-random string between the SEM server and each user in its domain. Typical example is the OTPs (one time passwords) achieved via secure ID tokens (e.g. The well known and widely used tamper- resistance RSA-Secured tokens [10]). The SEM and the user in the SEM's domain share a pseudo-random string which is updated every 30 (or 60) seconds at both parties. It is important to notice that this pseudo-random string is synchronously shared based on internal clocks implemented at both parties, consequently, the update is performed offline without any communication, hence this pseudo-random string cannot be reached via eaves- dropping. Let $X(\tau)$ be the pseudo-random string shared between Bob and the SEM at any given time interval,

VII. Deniability Transformation

O ur proposed scheme cannot withstand coercion of the sender, since a coerced sender is forced to reveal *R* and the index *i* which are verifiable by the coercer using the receiver's public key. A sender-deniable encryption is easily transformed to a receiver-deniable encryption and vice-versa. A sender-receiver deniable scheme requires n intermediaries, $I_1, ..., I_n$, with at least one of them remains honest (un attacked). The sender chooses n bits $b_1, ..., b_n$ such that L $b_i = b_t$ *and* sends *bi to* each I_i *using* the sender deniable public-key encryption

VIII. Bandwidth Improvement

It is possible to further improve the bandwidth of our receiver-deniable public-key encryption scheme as follows: Let $M = \{M_0, ..., M_{m-1}\}$ be the set of all possible strings According to the plaintext message, Alice sets the indices

$I = i_{v-1}, ..., i_0$ where *ij* points to *Mij* in *R*. In this case, each index *ij* is of ` bits where ` < lg lg

$N - \lg \grave{}$. The maximum number of indices per encryption (i.e. Contained in C_I) is v_{max} ' lg $N/\grave{}$. Since each index points to a string of ` bits, then, the encryption pair ($C_I$, $C_R$) encrypts about ` lg $N/\grave{}$ = lg N *bits* of plaintext. Hence, for a 1024 bits RSA modulus, a 2048 bits of cipher text encrypts 1024 − δ bits of plaintext where δ is the bit-length of *r*.

Finally, for each index, *ij* , Bob involves with the SEM in one invocation of OT1 oblivious transfer of strings to get *Mij* .

IX. Results and Discussion

We proposed a scheme for receiver-deniable public-key encryption. Our scheme is based on mediated RSA PKI. Our scheme proves efficiency over that proposed in [5] in the sense of bandwidth, deniability and decipherability. The scheme can be transformed to a sender-deniable or a sender-receiver-deniable using the tricks of [5]. The complexity of the oblivious transfer protocol used in our RD-PKE was studied and improved in [11]. The reader may have noticed that, our proposed scheme is not restricted to RSA. Our scheme could be applied to any PKI with the mediated property. A final thing worth noting is that when our receiver-deniable scheme is transformed to a sender-deniable one, it is no more a one-move scheme. To construct a one-move sender-deniable scheme, other approaches must be invented. For example, one may consider the sender-deniable scheme in [12].

References

[1] M. Bellare, and P. Rogaway, "Random oracles are practical: a paradigm for designing efficient protocols," *1st Conference on Computer and Communications Security*, pp.62-73, 1993.

[2] D. Boneh, X. Ding, G. Tsudik, and M. Wong, "A method for fast revocation of public key certificates and security capabilities," *Proceedings of the 10th USENIX Security Symposium*, pp. 297-308.

[3] C. Cachin, S. Micali, and M. Stadler, "Computationally private information retrieval with poly-logarithmic communication," *Advances in Cryptography - Eurocrypt '99*, pp. 402-414, 1999.

[4] R. Canetti, U. Feige, O. Goldreich, and M. Naor, "Adaptively secure multi-party computation," *Proceedings 28th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 639-648, 1996.

[5] R. Canetti, C. Dwork, M. Naor, and R. Ostrovsky, "Deniable encryption," *Proceedings of the 17th Annual international Cryptology Conference on Advances in Cryptology*, pp. 90-104, Springer-Verlag, London, 1997.

[6] R. Cramer, R. Gennaro, and B. Schoenmakers, "A secure and optimally efficient multi-authority election scheme," *Eurocrypt '97*, pp. 103-118, 1997.

[7] Y. Gertner, Y. Ishai, E. Kushilevitz, and T. Malkin, "Protecting data privacy in information retrieval schemes," *Proceedings of 30th Annual ACM Symposium on Theory of Computing*, pp. 151-160, 1998.

[8] S. Goldwasser, and S. Micali, "Probabilistic encryption," *Journal of Computer and System Sciences*, vol. 28, no. 2, pp. 270-299, 1984.

[9] M. Hirt, and K. Sako, "Efficient receipt-free voting based on homomorphic encryption," *Eurocrypt '00*, pp. 539-556, 2000.

[10] (http://www.rsa.com/rsalabs)

[11] M. H. Ibrahim, "Eliminating quadratic slowdown in two-prime RSA function sharing," *Int. Journal of Network Security (IJNS)*, vol. 7, no. 1, pp.107-114, 2008.

[12] M.H. Ibrahim, "A method for obtaining deniable public-key encryption," *International Journal of Network Security (IJNS)*, to appear.

IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 11, Issue 4 (May. - Jun. 2013), PP 08-12
www.iosrjournals.org

Deniable Encryption Key

P.Lokesh Kumar Reddy1[1], B.Rama Bhupal Reddy[2], S.Rama Krishna[3]
[1]*Rama Raja Institute of Technology and Science, Tirupati, A.P., India*
[2]*Dept. of Mathematics, K.S.R.M. College of Engineering, Kadapa, A.P., India*
[3]*Dept. of Computer Science, S.V. University, Tirupati, A.P., India*

Abstract: *Deniable encryption is an important that allows a user (a sender and/or a receiver) to escape a coercion attempted by a coercive adversary. Such an adversary approaches the coerced user after transmission forcing him to reveal all his random inputs used during encryption or decryption. Since traditional encryption schemes commit the user to his random inputs, the user is forced to reveal the true values of all his random inputs (including the encrypted/decrypted messages and the encryption/decryption keys) which are verifiable by this coercer using the intercepted cipher text. In this scenario, a coercer may force the user to perform actions against his wish. An appealing property in the mediated RSA, PKI was introduced that, the user has no information, neither about his full private (decryption) key, nor the factorization of the RSA public modulus, which represents an excellent step toward achieving in forcibility in public key encryption, since, a coercer cannot ask the user to reveal such unknown information. In this pa- per we present a scheme for receiver-deniable public-key encryption, by which, the receiver is able to lie about the decrypted message to a coercer and hence, escape a coercion. On one hand, the receiver is able to decrypt for the correct message, on the other hand, all the information held by the receiver, when opened to a coercer, do not allow this coercer to verify the encrypted message and consequently, approaching this user becomes useless from the very beginning.*

Keywords*: Deniable encryption, mediated PKI, oblivious transfer, public-key encryption, RSA*

I. Introduction

While traditional encryption schemes protect the privacy of the sender and the receiver against eavesdroppers (semantic security), they fail to provide protection against coercers. A coercive adversary has the power to approach the user (the sender and/or the receiver) after the ciphertext is transmitted and of course recorded by this adversary. It commands the user to reveal all his random inputs used during encryption or decryption. Since the cipher text produced, using standard encryption schemes (specially, public-key encryption) commits the user to his random inputs, It cannot lie about the true plaintext. Such commitments allow the coercive adversary to verify the validity of the opened message. Deniable encryption allows a user to escape coercion. Namely, if this user opens all his inputs (including the claimed encrypted message) to a coercer, the coercer fails to prove the validity or invalidity of the opened message.

Deniable encryption maybe classified according to which party is coerced: sender-deniable encryption schemes are resilient against coercing the sender. The Definitions for receiver-deniable and sender-receiver- deniable follow analogously. When the sender and the receiver initially share a common secret key, this is spoken off as shared-key deniable encryption. In deniable public- key encryption, no pre-shared information and no communications are assumed prior to the encryption process. This follows from the assumptions of standard public-key encryption schemes. Yet, deniable public-key encryption is more challenging than deniable shared-key encryption since the public key of the receiver is already known to everyone including the coercer, consequently, neither the sender nor the receiver can lie about the receiver's public key.

The work in [5] showed that it is possible by simple tricks to transform any sender-deniable encryption scheme to a receiver-deniable encryption scheme and vice- versa. Also, they showed that, with the help of other parties with at least one of them remains un-attacked, it is possible to transform a sender-deniable encryption scheme to a sender-receiver-deniable encryption scheme.

In our recent work of, we devised a sender-deniable public-key encryption based on quadratic residuosity of a composite modulus and showed how to device a sender- deniable public-key encryption from any trapdoor per mutation. However, when the schemes are transformed to be receiver-deniable using the tricks of [5], the schemes are no more one-move schemes. Considering only one- move schemes, receiver deniability is more challenging than sender-deniability since in the later case, everyone knows the public-key of the receiver but the private key of the receiver is known only to the receiver who is beyond the reach of the coercer. In the former case, the receiver maybe coerced to reveal his private key which is verifiable using the public key and a dummy message.

Deniable encryption is very useful in the protocols where coercive adversaries come to play as a potential threat. For example, deniable encryption protects voters from being coerced during electronic elections [6, 9]. It is also very useful to protect bidders in electronic auctions. Generally, deniable encryption is very important when a party is forced to act against his/her wish.

Our construction assumes the existence of the simple and efficient mediated-RSA (mRSA) [2] as the PKI in place. mRSA was invented as a method to achieve fast revocation in RSA PKI. mRSA involves a special entity, called the SEM (SEcurity Mediator), an on-line partially trusted server, to help signing or decrypting messages. The CA generates the private key d corresponding to Bob's (the receiver's) public key e and splits this private key into two pieces. One piece (d_{SEM}) is delivered to the SEM and the other piece (d_{Bob}) is delivered to Bob. The pair (e, N) is the usual RSA public key. An important property to notice here is that Bob himself has no information neither about his full private key, nor about the factorization of the public modulus N . This property is an excellent step toward achieving deniability since, a coercer will not ask Bob to reveal such unknown information. To decrypt a received cipher text, C, each party (Bob and SEM) performs his/her partial decryption on C; finally the partial decryptions are combined to recover the plaintext message M. To revoke Bob ability to sign or de- crypt messages, the CA instructs the SEM to stop issuing partial decryptions or signatures (spoken of as tokens) for Bob public key. At this instant, Bob's signature and/or decryption capabilities are revoked. The functionality is equivalent to (and indistinguishable from) standard RSA due to the fact that the splitting of the private key is transparent to the outside, i.e., to those who use the corresponding public key. Also, knowledge of a half-key cannot be used to derive the entire private key. Therefore, neither Bob nor the SEM can decrypt or sign a message without mutual consent.

As our PKI is established, we turn to discuss our tools. To complete the deniability service, we need an efficient protocol for (1-out-of-n) oblivious transfer (OT_1). Rabin proposed the concept of oblivious transfer (OT) in the cryptographic scenario. In this case the sender has only one secret bit b and would like to have the receiver to get it with probability ½, on the other hand, the receiver does not want the sender to know whether it gets b or not. For OT_1 , the sender has two secrets b_1 and b_2 , the receiver will get one of them at the receiver's choice.

The receiver does not want the sender to know which bit he chooses and the receiver must not know any information other than he chosen. Oblivious transfer is a fundamental primitive in many cryptographic applications and secure distributed computations and has many applications such as private information retrieval (PIR), fair electronic contract signing, oblivious secure computation, etc. Our proposed receiver-deniable public-key encryption scheme requires one invocation of OT_1 between Bob and the SEM. Also we need a Secured mechanism [10] to make the OT protocol deniable.

The rest of the paper is organized as follows: Section 2 describes the related work. Section 3 gives our motivations and contributions. The underlying PKI and the oblivious transfer protocol are described in Section 4.

Section 5 states our assumptions and model. . The strong version is given in Section 6. Section 7 shows the techniques to transform deniability. The bandwidth is improved in Section 8. Finally, the conclusions are given in Section 9.

II. Related Work

The work in [5] constructed a sender-deniable public- key encryption scheme based on trapdoor permutations. However, the scheme falls short of achieving an appropriate level of deniability, that is, to achieve a high deniability, the size of the cipher text corresponding to a one bit encryption is super-polynomial and hence inefficient. In the deniable public-key encryption scheme of [5], a one bit plaintext requires tn bits of cipher text where t is the bit-length of elements in a translucent set St and $t = s + k$ for security parameters n, s and k.

The scheme provides deniability of $4/n$ *and* decryption error of $n2^{-k}$. Hence, to achieve a high level of deniability and a sufficiently low decryption error, the cipher text is super- polynomial and almost impractical [5]. Constructed two deniable public-key encryption schemes based on translucent sets, the first represents the building block for the second which they have called, the "Parity Scheme". The work in [5] also notified that in order to build a one-round scheme, different approaches are required. Also, [5] introduced techniques for the less challenging, deniable shared- key encryption and showed that the one-time-pad is a perfect deniable shared-key encryption.

Based on the sender-deniable public-key encryption, the work in [4] described a general multiparty computations allowing a set of players to compute a common function of their inputs with the ability to escape a coercion. In fact, deniable encryption has an impact on designing adaptively secure multiparty computations [4] since, the notion of deniability is stronger than the notion of non-committing encryption Schemes.

III. Motivations and Contributions

In this Section we describe our motivations and contributions of the work in this paper.

3.1 Motivations

Deniable public-key encryption is a strong primitive, essential in all cryptographic protocols where a coercive adversary comes to play with high potential. Deniable public-key encryption realizes the "Receipt-freeness" attribute which is a very important attribute in electronic voting, electronic bidding and auctions. The schemes proposed in [5] fall short of achieving the desired level of deniability and correctness unless the size of the cipher text corresponding to a one bit encryption is super- polynomial. An appealing property in the mRSA PKI [2] is that the user himself has no information neither about his full private key, nor about the factorization of the public modulus N, consequently, a coercer will not ask the user for such unknown information.

3.2 Contributions

The contributions of this paper are to introduce an efficient receiver-deniable public-key encryption (RD-PKE) scheme. Our proposed scheme enjoys the following properties:

- It is a one-move scheme without any pre-encryption information required to be sent between the sender and the receiver prior to encryption.
- No pre-shared secret information is required between the sender and the receiver.
- Achieves a high level of deniability equivalent to the factorization of a large two-prime modulus.
- No deciphering errors.
- The bandwidth (cipher text bit-length) is significantly improved compared to previous constructions.

Efficiency, We reduce the required bandwidth (cipher- text bit-length) to, 2 lg N bits for a single bit encryption, where N is a two-prime RSA modulus. Moreover, this bandwidth can be efficiently improved, that is, 2 lg N bits of cipher text allow about lg $N - \delta$ bits of plaintext encryption where δ is a short randomizing string. At the same time, our scheme provides strong deniability (i.e. undetectable cheating) equivalent to the infeasibility to factor a sufficiently large two-prime modulus. Unlike the schemes of [5], our scheme produces no decryption errors and hence, more reliable. We introduce two versions of our RD-PKE scheme, a weak version to declare our idea and security proofs, and then we show a simple modification to improve this weak version to be a strong RD-PKE scheme.

IV. Preliminaries

4.1 Mediated RSA

Mediated RSA was invented as a simple method to achieve fast revocation in public-key cryptosystem. As usual, a trusted certificate authority (CA) sets up the RSA modulus N, the public exponent e and the private exponent d for the user. Next, instead of delivering d to the user, the CA splits d into two pieces d_{SEM} and d_{user} such that $d = d_{SEM} + d_{user} \bmod \phi(N)$ where $\phi(N)$ is the RSA Euler totient. Finally, the CA secretly delivers d_{user} *to* the user and d_{SEM} to the SEM.

Encryption, For Alice to encrypt a message M $\in Z_N$ to Bob, she uses Bob's public pair (N, e) to compute the usual RSA cipher text $C = M^e \bmod N$ and sends C to Bob.

Decryption, on the reception of C by Bob, the decryption process is as follows:

- Bob delivers C to the SEM.
- If Bob's key is revoked, the SEM returns ERROR and aborts, else,
- The SEM computes her partial decryption $P\ D_{SEM}$ = C $d_{SEM \bmod N}$ and returns P D_{SEM} to Bob.• Bob computes his partial decryption $P\ D_{Bob} = C\ d_{Bob}$

Mod N *and* extracts $M = P\ D_{SEM}\ P\ D_{Bob \bmod N}$.

It is important to notice that the SEM gains no information about the decrypted message M *[*2].

4.2. Oblivious Transfer

Our proposed RD-PKE requires that Bob involves with the SEM in an OT1 invocation to get his encrypted bit. The main objective of the oblivious transfer protocols was to improve the efficiency and security of the protocols. The protocols of [18] have several appealing properties.

First, they prove efficiency over previous protocols, second, there are no number theoretic constrains on the strings to be obliviously transferred, third, the protocols have bandwidth-computation tradeoffs which make them suitable for variety of applications. The protocols of operate over a group Z_q of prime order, more precisely, G_q is a subgroup of order q of Z_* where p is prime and $q/p - 1$. Let g be a generator group and assume that the Diffie-Hellman assumption holds. In their OT1: The sender owns two strings r_0 and r_1. He chooses a random element $U \in Z_q$ and publishes it. The chooser picks a random $1 \leq k \leq q$ and sets $pk\sigma = gk$ where$\in \{0, 1\}$ is the chooser's choice.

The chooser also computes pk1$-\sigma = U/pk\sigma$ and sends pk0 to the sender. The sender picks a random R and computes gR *and U R*, he also computes pk^R and $pk^R = U^R/pk^R$. The sender computes $pk_1-\sigma = U/pk\sigma$ and sends pk_0 to the sender. The sender picks a random R and computes gR *and U R*, he also computes pk_R and $pk_R = U^R/pk^R$. The sender The chooser selects a random k and sets $pk\sigma = gk$ where $\sigma \in \{0, ..., n-1\}$ is his choice, it holds that $pk^i = U^i/pk^0 \ \forall i = (1, ..., n-1)$. The chooser sends pk_0 to the sender. The sender computes pkR *as* well as pk^R

V. Assumptions and Model

We define a *receiver-deniable public-key encryption (RD- PKE) scheme* as a scheme by which, the receiver is able to lie about the decrypted message to a coercer and hence, escape coercion. On one hand, the receiver is able to decrypt for the correct message, on the other hand, all the information held (or extractable) by the receiver when opened to a coercer, do not allow this coercer to verify the encrypted message and consequently, approaching the receiver becomes useless from the very beginning.

The participants in our scheme are the certificate authority (CA), the security mediator (SEM), the sender (Alice), the receiver (Bob) and the coercive adversary (coercer). As usual, the CA is assumed to be fully trusted by all participants. The SEM is a semi-trusted party in the sense that it follows the execution steps word for word but it is willing to learn any information that could be leaked during execution. Alice is assumed to be beyond the reach of any coercer while Bob is possibly coerced.

The coercer has the power to approach Bob coercing him to reveal the decrypted message, the decryption partial key and all the parameters he used during decryption.

This paper describes the scheme allowing one bit encryption at a time. The reader will notice that the scheme can be easily adapted to allow multiple bits encryption at a time. We assume that an mRSA PKI is already in place. Hence, the pair (e, N) represents Bob's public key while d_{Bob} (respectively d_{SEM}) are the pieces of Bob's private key d held by Bob (respectively the SEM). Let *bt be* the true bit to be encrypted by Alice to Bob. The scheme is described next.

Encryption. To encrypt the bit *bt* to Bob, Alice proceeds as follows:

• Picks a lg *N bits* string R $\in_R ZN$. Let $r_0 ... r_{n-1}$ be the binary representation of *R*.
• Scans the binary representation of *R for* an index (Pointer) *i* such that $r_i = b_t$
• Computes and sends the two encryptions, $C_i = i \mod N$ *and* $C_R = R^e \mod N$ to Bob.

VI. Full Deniability

In this Section we show how to achieve full deniability in our RD-PKE scheme, i.e., the scheme will be deniable even if the coercer is capable of eavesdropping the Alice-Bob channel and the SEM-Bob channel as well. The problem is that the OT protocol is not deniable and hence commits Bob to what he receives from the SEM. We benefit from the fact that the SEM and all its users are in the same domain (or System) this fact facilitates the sharing of a time-synchronous pseudo-random string between the SEM server and each user in its domain. Typical example is the OTPs (one time passwords) achieved via secure ID tokens (e.g. The well known and widely used tamper- resistance RSA-Secured tokens [10]). The SEM and the user in the SEM's domain share a pseudo-random string which is updated every 30 (or 60) seconds at both parties. It is important to notice that this pseudo-random string is synchronously shared based on internal clocks implemented at both parties, consequently, the update is performed offline without any communication, hence this pseudo-random string cannot be reached via eaves- dropping. Let $X(\tau)$ be the pseudo-random string shared between Bob and the SEM at any given time interval,

VII. Deniability Transformation

O ur proposed scheme cannot withstand coercion of the sender, since a coerced sender is forced to reveal *R* and the index *i* which are verifiable by the coercer using the receiver's public key. A sender-deniable encryption is easily transformed to a receiver-deniable encryption and vice-versa. A sender-receiver deniable scheme requires n intermediaries, $I_1, ..., I_n$, with at least one of them remains honest (un attacked). The sender chooses n bits $b_1, ..., b_n$ such that L $b_i = b_t$ *and* sends *bi to* each I_i *using* the sender deniable public-key encryption

VIII. Bandwidth Improvement

It is possible to further improve the bandwidth of our receiver-deniable public-key encryption scheme as follows: Let $M = \{M_0, ..., M_{m-1}\}$ be the set of all possible strings According to the plaintext message, Alice sets the indices

$I = i_{v-1}, ..., i_0$ where *ij* points to *Mij* in *R*. In this case, each index *ij* is of ` bits where ` < lg lg

$N - \lg \grave{}$. The maximum number of indices per encryption (i.e. Contained in C_I) is v_{max} ' lg $N/\grave{}$. Since each index points to a string of ` bits, then, the encryption pair ($C_I$, $C_R$) encrypts about ` lg $N/\grave{}$ = lg N *bits* of plaintext. Hence, for a 1024 bits RSA modulus, a 2048 bits of cipher text encrypts 1024 − δ bits of plaintext where δ is the bit-length of *r*.

Finally, for each index, *ij* , Bob involves with the SEM in one invocation of OT1 oblivious transfer of strings to get *Mij* .

IX. Results and Discussion

We proposed a scheme for receiver-deniable public-key encryption. Our scheme is based on mediated RSA PKI. Our scheme proves efficiency over that proposed in [5] in the sense of bandwidth, deniability and decipherability. The scheme can be transformed to a sender-deniable or a sender-receiver-deniable using the tricks of [5]. The complexity of the oblivious transfer protocol used in our RD-PKE was studied and improved in [11]. The reader may have noticed that, our proposed scheme is not restricted to RSA. Our scheme could be applied to any PKI with the mediated property. A final thing worth noting is that when our receiver-deniable scheme is transformed to a sender-deniable one, it is no more a one-move scheme. To construct a one-move sender-deniable scheme, other approaches must be invented. For example, one may consider the sender-deniable scheme in [12].

References

[1] M. Bellare, and P. Rogaway, "Random oracles are practical: a paradigm for designing efficient protocols," *1st Conference on Computer and Communications Security*, pp.62-73, 1993.

[2] D. Boneh, X. Ding, G. Tsudik, and M. Wong, "A method for fast revocation of public key certificates and security capabilities," *Proceedings of the 10th USENIX Security Symposium*, pp. 297-308.

[3] C. Cachin, S. Micali, and M. Stadler, "Computationally private information retrieval with poly-logarithmic communication," *Advances in Cryptography - Eurocrypt '99*, pp. 402-414, 1999.

[4] R. Canetti, U. Feige, O. Goldreich, and M. Naor, "Adaptively secure multi-party computation," *Proceedings 28th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 639-648, 1996.

[5] R. Canetti, C. Dwork, M. Naor, and R. Ostrovsky, "Deniable encryption," *Proceedings of the 17th Annual international Cryptology Conference on Advances in Cryptology*, pp. 90-104, Springer-Verlag, London, 1997.

[6] R. Cramer, R. Gennaro, and B. Schoenmakers, "A secure and optimally efficient multi-authority election scheme," *Eurocrypt '97*, pp. 103-118, 1997.

[7] Y. Gertner, Y. Ishai, E. Kushilevitz, and T. Malkin, "Protecting data privacy in information retrieval schemes," *Proceedings of 30th Annual ACM Symposium on Theory of Computing*, pp. 151-160, 1998.

[8] S. Goldwasser, and S. Micali, "Probabilistic encryption," *Journal of Computer and System Sciences*, vol. 28, no. 2, pp. 270-299, 1984.

[9] M. Hirt, and K. Sako, "Efficient receipt-free voting based on homomorphic encryption," *Eurocrypt '00*, pp. 539-556, 2000.

[10] (http://www.rsa.com/rsalabs)

[11] M. H. Ibrahim, "Eliminating quadratic slowdown in two-prime RSA function sharing," *Int. Journal of Network Security (IJNS)*, vol. 7, no. 1, pp.107-114, 2008.

[12] M.H. Ibrahim, "A method for obtaining deniable public-key encryption," *International Journal of Network Security (IJNS)*, to appear.

IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 11, Issue 4 (May. - Jun. 2013), PP 13-19
www.iosrjournals.org

E-Healthcare Billing and Record Management Information System using Android with Cloud

Vanitha T N[1], Narasimha Murthy M S[2], Chaitra B[3]
[1](Fourth Semester M.Tech, Department of Computer Science & Engineering, Acharya Institute of Technology, Bangalore, India)
[2](Assistant Professor, Department of Computer Science & Engineering, Acharya Institute of Technology, Bangalore, India)
[3](Assistant Professor, Department of Computer Science & Engineering, Acharya Institute of Technology, Bangalore, India)

Abstract : *In today's civilized society, the people are betrayed with proper healthcare facilities. In order to minimize the cost and complexity involved in processing traditional billing system, Electronic health records (EHR) and electronic billing systems have been proposed as mechanisms to help curb the rising costs of healthcare and also helps to detect the fraudulent practices in healthcare system. The introduction of Cloud computing concept in electronic healthcare systems is the solution for better utilization of healthcare facilities. It uses open-source cloud computing technologies as the mechanism to build an affordable, secure, and scalable platform that supports billing as well as EHR operations. We call this platform as "MedBook" which is a cloud solution that provides patients, healthcare providers, and healthcare payers a platform for exchange of information about EHR, billing activities, and benefits inquiries. MedBook serves as an integration point between the various participants in the healthcare delivery system. This paper presents the architecture and implementation status of this system. The developed system has been evaluated using the Jelastic cloud service. MedBook is a Software-as-a-Service (SaaS) application built on top of open source cloud technologies and running on an Infrastructure-as-a-Service (IaaS) platform. The client applications are mobile apps run from Google's Android enabled devices.*

Keywords - *Cloud Computing, EHR, Mobile apps, Open source Cloud, REST-based API, SaaS*

I. INTRODUCTION

Electronic health records (EHR) and electronic billing systems have been proposed as mechanisms to curb the rising costs of healthcare and also helps to detect the fraudulent practices in the traditional healthcare system[1]. Many healthcare professionals, hospitals and insurance agencies maintains the paper-based records, billing of the patients which is been converted later into computer-based billing and records which can be abused, modified or lost for malpractice done by frauds either for money or grudge.

Hence the personal information of the patients is revealed, bogus information are entered and misused in traditional Healthcare system. Moreover Traditional healthcare system depends on the centralized server which is unreliable, insecure in accessing, storing medical data regardless of time, cost and location. Hence it is more complex and lack privacy and cost involved in integrating medical information is expensive.

Given this scenario, Electronic Health Records (EHR) and Electronic Medical Billing (EMB) have been proposed as a mechanism which reduces healthcare disparities and ensures adequate privacy and security. One potential solution for addressing all aforementioned issues is the introduction of Cloud Computing concept in electronic healthcare systems. This mechanism pursued the idea of using open-source public cloud computing Technologies and mobile plus cloud paradigm [2] to build an affordable, secure and scalable platform that supports billing as well as EHR operations. We call this platform as *MedBook* and in this paper we present the proposed architecture and implementation status of this system. MedBook is a cloud solution that provides patients, healthcare professionals/providers and healthcare payers a platform for exchange of electronic information about billing activities, benefit inquiries and EHR operations such as insert delete and update record using open source cloud services and Android operating system(OS).

MedBook is Software–as-a-Service (SaaS) platform built on top of open source public cloud technologies and running on the top of an Infrastructure-as-a Service (IaaS) platform [3]. Generally the server applications are implemented as a collection of web services and web applications using MySQL, Tomcat 6or7 server, Apache web server. All the web services run on virtual machines powered by Windows XP or Ubuntu

Linux 10.04.These servers are hosted inside an open source cloud [4] which can be Jelastic, Eucalyptus 2.0, Open Stack and so forth.
The client applications are mobile apps run from Google's Android Enabled phones [5]. These client applications are built using Java 1.7 or 1.6 and uses REST based API to interact with MedBook SaaS Infrastructure.

The rest of the paper is organized as follows. Section II provides the basics of cloud computing and EHR details. Section III provides the Motivation for the use of MedBook. Section IV presents the proposed system architecture of MedBook SaaS application and its implementation status. Section V discusses the scope of the work. Finally, section VI presents a summary and conclusion of the paper.

II. CLOUD COMPUTING AND ELECTRONIC HEALTH INFORMATION

According to NIST " Cloud computing is a pay per use model for enabling convenient, on demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction[6].

Cloud Computing provides the following Key Characteristics [7]:

(a)Broad network access Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, laptops, and PDAs).
(b)Resource pooling The provider's computing resources is pooled to serve multiple consumers using a *multi-tenant* model; *Multi-tenancy* [8] enables sharing of resources and costs across a large pool of users thus allowing for:

- *Centralization* of infrastructure in locations with lower costs (such as real estate, electricity, etc.)
- *Peak-load capacity* increases (users need not engineer for highest possible load-levels)
- *Utilisation and efficiency* improvements for systems that are often only 10–20% utilised.

Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.

(c)Application programming interface (API) accessibility to software that enables machines to interact with cloud software in the same way the user interface facilitates interaction between humans and computers. Cloud computing systems typically use REST-based APIs [9].
(d)Cost is claimed to be reduced when using cloud infrastructure.
(e)Reliability is improved if multiple redundant sites are used, which makes well-designed cloud computing suitable for business continuity and disaster recovery.
(f)Security could improve due to centralization of data, increased security-focused resources, etc., but concerns can persist about loss of control over certain sensitive data, and the lack of security for stored kernel.

Cloud computing has three service models to deliver various services namely, *Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS)* [7].

Software as a Service (SaaS) In this Service model, an application software is hosted in the cloud and thus the users can access the application through Mobile apps, emails, PC's. The users need not have to control the cloud infrastructure since the model itself employs the multi-tenancy system architecture for accessing the application and provides optimization in terms of speed, security, availability, disaster recovery and maintenance. Example of SaaS includes Google Mail, Google Docs, Salesforce.com, ZOHO and so forth.
Platform as a Service (PaaS) In this Service model, the users has to deploy onto the cloud infrastructure created using programming languages and tools and configuration management supported by the provider. The users need not have to control the cloud infrastructure including network, servers, operating systems, or storage. PaaS model offers a development platform to host both the completed and in- progress cloud application. Example of PaaS includes Google AppEngine, AWS.
Infrastructure as a Service (IaaS) In this Service model, the users directly use processing, storage, networks, and other computing resources provided in the IaaS cloud which employs Virtualization in order to integrate/decompose physical resources as per user requirements. The users need not have to manage or control the cloud infrastructure but has control over operating systems, storage, deployed applications. Example of IaaS includes Amazon's EC2, Sun Microsystems.

Cloud computing provides various deployment models they are given as follows, *Private cloud, Public cloud, Hybrid cloud and community cloud* [7].

Private cloud The cloud infrastructure is operated within a single organization, and it is managed by the organization itself or by a third-party which may exist on/ off premises.

Public cloud The cloud infrastructure is made available to the general public or a large industry group and is owned by service providers. These services are free or offered on a pay-per-use model. Generally, public cloud service providers like Amazon EC2, S3, Microsoft and Google AppEngine.

Community cloud The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party and may exist on/off premises.

Hybrid cloud The cloud Infrastructure is a composition of two or more clouds (private, community or public) that remain unique entities but are bound together, offering the benefits of multiple deployment models.,

Given the characteristics of Cloud Computing and its service models, thus improves the ability of accessing the information by the users being able to rapidly and inexpensively re-provision technological infrastructure resources. Device and location independence enable users to access systems using a web browser regardless of their location or what device they are using (e.g., mobile phones). Multi-tenancy enables sharing of resources and costs across a large pool of users thus allowing for centralization of infrastructure in locations with lower costs. Reliability improves through the use of multiple redundant sites, which makes Cloud Computing suitable for business continuity and disaster recovery. Security typically improves due to centralization of data and increased security-focused resources. Sustainability comes about through improved resource utilization, more efficient systems.

Electronic health information is accessed by all the participants of healthcare system such as patients, healthcare providers, healthcare payer using open source cloud which acts as a server that faces several challenges, like data storage and management (e.g., physical storage issues, availability and maintenance), interoperability and availability of heterogeneous resources, security and privacy (e.g., permission control, data anonymity, etc.), unified and ubiquitous access.

The mobile apps such as Google's Android operating system is used as a client which focus towards achieving two specific goals [10]: the availability of e-health applications and medical information anywhere and anytime and the invisibility of computing. Mobile apps basically support electronic billing and EHR activities of patient and their medical history which can be accessed individually by patient, healthcare provider, healthcare payer by authenticating themselves with MedBook cloud server.

III. MOTIVATION

MedBook provides a highly reliable and secure electronic billing and record management system. It also helps reduce the occurrences of medical errors due to incomplete medical information. Privacy of medical information is maintained to prevent unauthorized access and misuse of electronic information [11].Since the MedBook is Mobile plus cloud paradigm, the various participants such as patients, health care payers, healthcare professional can exchange information regardless of time, location, cost involved in it. Since the proposed system utilizes open source cloud computing technologies, the interaction of healthcare participants with MedBook SaaS application can be done globally which reduce the cost associated in accessing medical information to certain limit [12]. In addition, by integrating and correlating the billing system with EHR, it becomes possible to find that a given procedure was actually performed or the medical history of the patient utilized such procedures.

IV. Proposed System Architecture and Implementation Details

Fig. 1 Depicts the MedBook proposed system architecture. MedBook SaaS application serves as an integration point between the various participants in the healthcare delivery system such as Patients, healthcare providers, healthcare payer [3]. MedBook architecture basically contains two Modules such as *Client Module* which uses the Mobile apps such as android enabled phones to interact with MedBook application. *Server Module* which consists of a series of web services and databases residing inside an IaaS cloud that maintains the information about each patient's EHR.

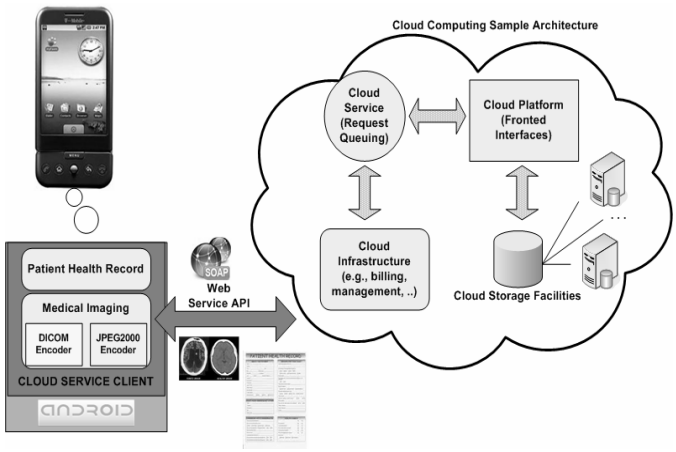


Fig. 1 MedBook proposed system architecture

A. MedBook Client Application

The MedBook Client applications are mobile apps which uses Android enabled phones that connect with the MedBook SaaS infrastructure by means of REST based API. The architecture of MedBook Client application is depicted in Fig. 2. An android enabled phone is an open source and basically supports large number of applications compared to other Smart phones [13]. Android client application is designed for Patients, Healthcare provider/professional, Healthcare payers to perform billing and EHR activities with MedBook.

The Android client application for patient is designed to perform following operations to:

- Access Medical details prescribed
- Know Insurance Benefit plan
- Access their EHR Information

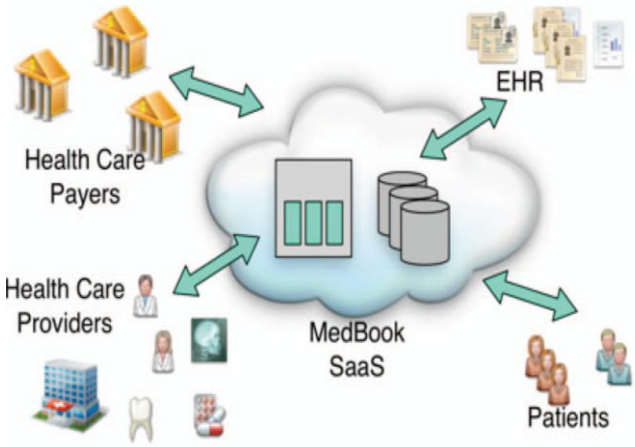


Fig. 2 MedBook Client Architecture

The Android client application for Healthcare payers is designed to perform following operations:

- To receive billing request
- Response to Benefit and Enrollment verification request
- Response to authorization request for procedures/drugs
- To enter Insurance Benefit plan details of a particular patient

The Android client application for Healthcare provider/professional is designed to perform following operations to:

- Submit billing Claims
- Get status information on claims
- Manage EHR of the patient i.e., access and update

- Request authorization for treatment/drugs
- Enter patient medical details
- Submit benefit enquiries

B. MedBook Server Application

MedBook consists of a series of web services and databases residing inside an IaaS cloud that maintains the information related to each patient's EHR activities [14]. The architecture of MedBook server application is depicted in Fig. 3. The server module is designed to perform following activities such as

- Service description
- Submitting Billing Transactions
- Representing and accessing EHR's

Service description The description of the services used in the MedBook SaaS application is as follows [3]:

(a)Billing Collector This web service receives the list of billing requests from the healthcare professionals on behalf of patient, and sends to the billing service instances for validation and further processing done by healthcare payers.

(b)*Billing Service* This web service is an interface between the healthcare payers with respect to healthcare provider and patient. It performs the verification operation on the billing request and reports back to healthcare provider about errors or omissions. Successfully validated claims are sent to the healthcare payer. This service also receives notifications from the healthcare payer (through payer gateway service) regarding the queries related to Patient medical history. It also provides notifications to the healthcare providers or patient which can be fetched by the client application.

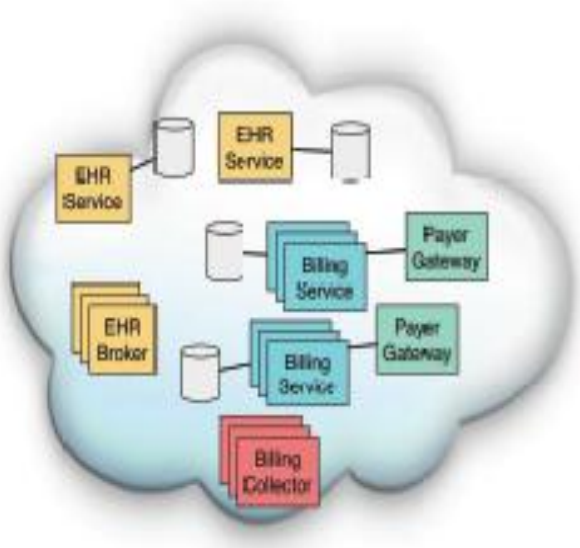


Fig. 3 MedBook Server Architecture

(c)EHR broker This web service takes care of locating all the patient medical history through EHR. Since each EHR is given a unique Patient identification number (PIN) to identify the EHR of the patient and a copy of the PIN. This EHR broker can be contacted by the client application or by the other web services such as the billing service or the payer gateway.

(d)EHR Service This web service provides read-write-update operations for a collection of the EHR segments. EHR service communicates securely by means of encrypted HTTP channels.

(e)Payer Gateway This web service provides an Interface to the healthcare payer to communicate with other services. This service translates all the notifications and request made by the MedBook to native formats used by the healthcare payer. Hence the payer gateway provides a common set of operations and vocabulary to be used between MedBook and the healthcare payers to perform this translation.

Submitting Billing Transactions Typically electronic transactions for medical domains can be classified for the following purpose such as

(a)Enrollment and Benefits Verification used to verify the patient's EHR in a health plan and check whether the patient is valid/Invalid for accessing the benefits of healthcare plan by the healthcare payer from patient's EHR.

(b)Healthcare Claim used to submit the actual bill of each patient from healthcare provider to healthcare payer for the prescriptions or procedures used in the treatment of patient.
(c)Healthcare Claim status and Notification used to verify the authorization request for procedures is submitted successfully or not to healthcare payer and also receives notification regarding healthcare claim that the billing request is valid or invalid by healthcare payer to the healthcare provider.

Representing and accessing EHR's In MedBook system, each patient has a unique identification number called the Patient Identification Number (PIN) [14] that is used to uniquely identify their EHR. MedBook admin has only the rights to delete a particular EHR when not necessary.EHR is considered as a validation tool against billing transactions submitted for medical procedures. Billing systems should be used with EHR systems because the procedure used by the patient becomes the part of the claims and these data should reflect in patient's EHR also.

Thus the use of EHR has strived for simplicity, scalability and security purpose. In our MedBook system, the portion of the EHR is called Segment and they are stored in a collection of cloud-resident relational database instances. The healthcare provider in MedBook system has control over the segment to render services to patient and the patient manages the demographic information as well as personal information through unique ID and password. Scheme is depicted in Fig. 4. Whenever the healthcare provider adds a new segment into the EHR of a patient two actions are performed firstly the entry with the information about conditions, diagnosis and treatments are stored in the database [15]. This is done through the EHR service and thus EHR is effectively updated to the local segment of the database. Secondly EHR service sends a message to the EHR broker to add a new entry to the segment indexing service which includes the form: (PIN, segment id, database id). Here the database id is a unique id for the database instance that stores the new segment created by the healthcare provider. The EHR broker then delivers all these entries to the client application for display or analysis purpose.

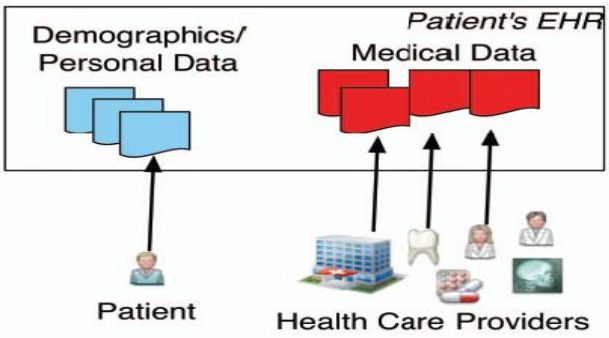


Fig. 4 EHR Control Split between healthcare participants

In our MedBook system, billing process starts when the Billing Collector receives one or more billing transactions. The Collector accepts claims from healthcare providers either as individual transactions or in a batch/queue system. Billing service instance takes the billing job out from the queue system and begins working on each claim. The first task of the billing Service is to verify that the claim is valid or invalid and stores them in a database which the healthcare provider can fetch and correct them for resubmission. Successfully validated claims are also stored into a database and submitted to a Payer Gateway which is then translated into the format used by the internal billing systems to be used by the healthcare payer.

C. Implementation status

All the web services were implemented following the Model-View-Controller (MVC) pattern. We used XML to write and deploy the services. All storage is provided using MySQL. All the web services are run on virtual machines powered by Windows XP and all these machines are hosted within an open source cloud namely Jelastic cloud[16][17]. The client applications are Mobile apps such as Android 3.0 and above enabled phones.

D. Utilizing Jelastic Cloud Service

For the realization of the mobile healthcare information management system, the Jelastic open source cloud service has been utilized [17]. The main reason for selecting the specific Cloud Computing platform is that it is a commercial service well established and used successfully in several applications. It provides users with several interoperable web interfaces for managing data (SaaS model) and developers with the ability to create their own applications for accessing the latter (IaaS model) and is suitable for managing healthcare information . It provides the technologies necessary to build both private and public clouds, with a goal of standardizing cloud computing infrastructure.

V. Scope of Work

The proposed work helps to curb the rising cost of healthcare to certain limit using open source cloud technologies [18]. Our MedBook system provides solutions to certain top issues in healthcare sector. Firstly, Lack of Information exchange between medical practitioners. Secondly, privacy of Information to prevent unauthorized disclosure of information. Thirdly, Security of IT systems to prevent healthcare data being stolen or intentionally corrupted. Fourthly, accuracy of data to prevent incorrect information to be entered into the EHR.

VI. Conclusion

Electronic health records (EHR) and Electronic Billing Systems have been proposed as mechanisms to help curb the rising costs of health care system. Given this scenario, the paper discusses the idea of using open-source cloud computing technologies as a cloud solution to build an affordable, secure, and scalable platform that supports billing as well as EHR operations which utilizes Jelastic open source cloud storage service. We call this platform as MedBook, and in this paper we presented the architecture and implementation status of this system. MedBook serves as an integration point between the various participants such as patients, health care providers, and health care payers a platform for exchange of information about EHR, billing activities, and benefits inquiries in the healthcare delivery system.

The proposed system namely, MedBook SaaS application uses open source public cloud services, hence the participants of the healthcare system can access information and relevant healthcare facilities regardless of time, location, cost and privacy factors involved in it. The sharing of medical information resources (electronic health data and corresponding processing applications) is a key factor playing an important role towards the successful adoption of Cloud Computing with Android enabled phones in healthcare systems.

References

[1] A.S. Boranbayev, and S.N. Boranbayev, "Development and Optimization of Information Systems for Health Insurance Billing", 2010 *IEEE 7th International Conference on Information Technology: New Generation, pp- 606 - 613, dated: 12-14 April 2010, Las Vegas, NV.*

[2] AM.Rodriguez et.al."Open911: Experiences with the Mobile Plus Cloud Paradigm", in Proc.2011 *IEEE Cloud Computing Conference, dated: 4-9 July 2011, Washington, DC,USA.*

[3] Manuel Rodriguez-Martinez, Harold Valdivia, Jose Rivera, Jaime Seguel, Melvin Greer, "MedBook: A Cloud-based Healthcare Billing and Record Management System", *2012 IEEE Fifth International Conference on Cloud Computing, pp- 899 – 905, dated: 24-29 June 2012 , DOI 10.1109/CLOUD.2012.133.*

[4] Joe Brockmeier, community evangelist for Cloud Stack Citrix" *What is an Open source cloud?"* released on March 28,2013.URL: http://www.linux.com/news/featured-blogs/196-zonker/711498-what-is-open-source-cloud-

[5] "Android Overview" Open Handset Alliance, retrieved on: 15-02-2012.URL:http://www.openhandsetalliance.com/

[6] P. Mell and T. Grance, *"Draft NIST working definition of cloud computing - v15,"* National Institutes of Standard and Technology, Information Technology Laboratory, dated: 19th, August-*2009.*

[7] Tharam Dillon, Chen Wu and Elizabeth hang,"Cloud computing; Issues and Challenges", *AINA, pp.27-33, dated: 20-23, April- 2010, 24th IEEE International Conference, Perth, WA.*

[8] A white paper from Juniper Networks Inc on "*Securing Multitenancy and cloud Computing*", Juniper Networks Inc 2012.pp.1-5,URL: http://www.juniper.net/us/en/local/pdf/whitepapers/2000381-en.pdf

[9] Ben Ramsey "*Designing RESTful Web applications*" released on September 13, 2007. http://files.benramsey.com/talks/2007/phpworks/phpworks07-rest.pdf

[10] Alvin Ybanez *"Best Android apps for personalizing and customizing your phone"* release on 2012-07-13. Retrieved on 09-11-2012. URL: http://www.androidauthority.com/best-apps-customizing-personalizing-android-phones-100685

[11] L. S. Liu, P. C. Shih, G. R. Hayes, "Barriers to the adoption and use of personal health record systems", in Proc. 2011 ACM conference, New York, NY.

[12] S. M. Thompson, and M. D. Dean, "Advancing information Technology in health care", Communications of the ACM, vol. 52, no. 6, 2009.

[13] Palo Alto, Singapore and Reading (UK)*"Google's Android becomes the world's leading smart phone platform"* Canalys Research release on. January 31, 2011. Retrieved on 15-02-2012.URL:http://www.canalys.com/newsroom/googles-android-becomes-worlds-leading-smart-phone-platform

[14] N. Stafford, "*Who owns the data in an Electronic Health Record*?" EHR Institute, Retrieved on February 29, 2012.URL: http://www.ehrinstitute.org/articles.lib/items/who-owns-the-data-in

[15] R. Green, "*Obstacles to EHR Adoption Lie in Small Group Practices*". URL: https://www.gplus.com/telecommunicationsservices/Insight/obstacles-to-ehr-adoption-lie-in-small-group practices- 38092, retrieved on: February 28, 2012.

[16] Why Jelastic? Because it's easy! http://jelastic.com/why

[17] "Jelastic cloud hosting", URL: http://www.layershift.com/hosting-services/jelastic-cloud-hosting

[18] Sebastian Rupley "*Top Open source resources for cloud computing"* released on Nov 6, 2009. URL: http://gigaom.com/2009/11/06/10-top-open-source-resources-for-cloud-computing

IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 11, Issue 4 (May. - Jun. 2013), PP 27-36
www.iosrjournals.org

Performance Analysis of Rician Fading Channels using Non-linear Modulation Methods with Memory Schemes in Simulink environment

Sunil Kumar.P[#1], Dr.M.G.Sumithra[#2], Ms.M.Sarumathi[#3]
[#1] *P.G.Scholar, Department of ECE, Bannari Amman Institute of Technology, Sathyamangalam*
[#2] *Professor, Department of ECE, Bannari Amman Institute of Technology, Sathyamangalam*
[#3] *Assistant Professor, Department of ECE, Bannari Amman Institute of Technology, Sathyamangalam*

Abstract : *When a signal travels from transmitter to receiver over multiple reflective paths then the phenomenon is called as multipath propagation. In a wireless mobile communication system, the multipath propagation can cause the degradation in the received signal strength and this process is called as fading. When a strong stationary path such as a line of sight path is introduced into the Rayleigh fading environment, the fading becomes Rice-distributed fading. Ricean fading is utilized for characterizing satellite communications and in some urban environments. In this paper, the performance analysis of Ricean Fading Channels using Non-linear modulation methods with memory Schemes is implemented and the Bit Error Rate is analyzed using Simulink tool.*

Keywords - *Fading, Ricean, Non-linear, Memory, Simulink*

I. INTRODUCTION TO DIFFERENT TYPES OF CHANNEL FADING

There are usually three types of channel fading for mobile communications: shadowing (slow fading), multipath Rayleigh fading, and frequency-selective fading. Reflection, diffraction, and scattering are the three major mechanisms that influence the signal propagation.

1.1. Log-Normal Shadowing:

During the motion of an MS, clutter such as trees, buildings, and moving vehicles partially block and reflect the signal, thus resulting in a drop in the received power. In the frequency domain, there is a power decrease in a wide frequency range. Hence, it is called slow fading. The slow power variation relative to the average can be modelled by a log-normal probability function (pdf). For the log-normal distribution, the logarithm of the random variable has a normal distribution. The pdf and cumulative distribution function (cdf) are given by

$$\rho(r) = \frac{1}{r\sigma\sqrt{2\pi}} e^{-\frac{(\ln r - m)^2}{2\sigma^2}} \tag{1}$$

where m and σ are the mean and deviation, $P_r(.)$ is the probability function, and erf (x) is the well-known error function. In the shadowing model, the transmit-to-receive power ratio $\psi = \frac{P_t}{P_r}$ is assumed to be random with a log-normal distribution [1] and is expressed as follows

$$\rho(\psi) = \frac{10/\ln 10}{\sqrt{2\pi}\sigma_{\psi dB}\psi} \exp\left(-\frac{(\psi_{dB} - m_{\psi dB})^2}{2\sigma^2_{\psi dB}}\right), \psi > 0 \tag{2}$$

where $\psi_{dB} = 10\log_{10}\psi$ in decibels, and $m_{\psi dB,}\sigma_{\psi dB}$ are the mean and standard deviation of $\psi_{dB.}$. In this model, it is possible for ψ to t2ake on a value within 0 and 1, which corresponds to $P_r > P_{t,}$ but this is physically impossible. Nevertheless, this probability is very small when $m_{\psi dB}$ is a large and positive number. This fluctuation in mean power occurs on a large scale, typically dozens or hundreds of wavelengths, and thus, is also known as large-scale fading or macroscopic fading. Statistically, macroscopic fading is determined by the local mean of a fast fading signal. Trees cause an important class of environmental clutter. A tree with heavy foliage causes shadowing. A tree with full foliage in summer has approximately a 10 dB higher loss due to shadowing than the same tree without leaves in winter as it acts as a wave diffractor.

1.2. Rayleigh Fading:

When both the I and Q components of the received signal, x_I and x_Q normally distributed, the received signal is a complex Gaussian variable. The envelope of the received signal, $r = (x_I^2 + x_Q^2)^{1/2}$, is Rayleigh distributed, and r^2 has an exponential distribution. Note that the exponential distribution is a special case of the central- χ^2 distribution with m=1. The χ^2 distribution is the distribution of $Y=X^2$, where X is a Gaussian distribution. Assuming both x_I and x_Q have a standard deviation of σ, the total power in the received signal is $E[r^2]/2=\sigma^2$, and the pdfs is expressed by the following mathematical expression

$$\rho_r(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}, 0 \le r < \infty \quad \text{and} \quad \rho_r{}^2(r) = \frac{1}{2\sigma^2} e^{-\frac{r}{2\sigma^2}} \tag{3}$$

For wireless communications, the envelope of the received carrier signal is Rayleigh distributed; such a type of fading is thus called Rayleigh fading. This can be caused by multipath with or without the Doppler effect. In the multipath case, when the dominant signal becomes weaker, such as in the non LOS case, the received signal is the sum of many components that are reflected from the surroundings. These independent scattered signal components have different amplitudes and phases (time delays); then the I and Q components of the received signal can be assumed to be independent zero-mean Gaussian processes. This is derived from the central limit theorem, which states that the sum of a sufficient number of random variables approaches very closely to a normal distribution.

1.3. Ricean Fading:

When a strong stationary path such as a LOS path is introduced into the Rayleigh fading environment, the fading becomes Rice-distributed fading. Ricean fading is suitable for characterizing satellite communications or in some urban environments. Ricean fading is also a small-scale fading. In this case, the probability of deep fades is much smaller than that in the Rayleigh-fading case. Based on the central limit theorem, the joint pdf of amplitude r and phase ϕ may be expressed as the following equation [2]

$$\rho_{r,\phi}(r,\phi) = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2+A^2-2rA\cos\phi}{2\sigma^2}} \tag{4}$$

where A is the amplitude of the dominant component and σ is the same as that for Rayleigh fading,. This joint pdf is not separable, and the pdf of r or ϕ can be obtained by integrating over the other quantity. The pdf of the amplitude is a Rice distribution and is mathematically expressed as follows [3]

$$\rho_r(r) = \frac{r}{\sigma^2} e^{-\frac{r+A^2}{2\sigma^2}} I_0\left(\frac{rA}{\sigma^2}\right), 0 \le r < \infty \tag{5}$$

where $Z_0(x)$ is the modified Bessel function of the first kind and zero order, and is defined as follows

$$Z_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{-x\cos\theta} d\theta \tag{6}$$

The mean square value of r is given by

$$\rho_r = 2\sigma^2 + A^2 \tag{7}$$

The Rice factor K*r* is defined as the ratio of the dominant component to the power in all the other components and it is given by the equation $K_r = \frac{A^2}{2\sigma^2}$ [4].The Rice distribution approximates the Rayleigh distribution with mean value A as Kr<<1, and reduces to it at Kr=0. It approximates the Gaussian distribution with mean value A as Kr>>1, and reduces to the Gaussian as $K_r \to \infty$.The factor Kr typically shows an exponential decrease with range, and varies from 20 near the BS to zero at a large distance. The dominant component changes the phase distribution from the uniformly random distribution of Rayleigh fading to clustering around the phase of the dominant component. The stronger the dominant component, the closer the resulting phase to the phase of the dominant component. This is similar to a delta function. Flat Ricean fading channel is suitable for characterizing a real satellite link.

1.4. Nakagami Fading:

The Nakagami distribution is another popular empirical fading model [5,6]

$$\rho(r) = \frac{2}{\Gamma(m)}(\frac{m}{2\sigma^2})^m r^{2m-1} e^{-m\frac{r^2}{2\sigma^2}}, r \geq 0 \quad (8)$$

Where $\sigma^2 = \frac{1}{2}E[r^2], \Gamma(.)$ is the Gamma function, and m $\geq 1/2$ is the fading figure. The received instantaneous power r^2 satisfies a Gamma distribution. The phase of the signal is uniformly distributed in [0,2π], which is independent of r. The Nakagami distribution is a general model obtained from experimental data fitting. The Nakagami distribution has a shape very similar to that of the Rice distribution. The shape parameter m controls the severity of fading. When m=1, the fading is close to Ricean fading, and the Nakagami and Rice distributions can approximate each other with the following mathematical expression,

$$K_r = (m-1) + \sqrt{m(m-1)}, m > 1 \quad (9)$$

$$m = \frac{(K_r+1)^2}{2K_r+1} \quad (10)$$

The Nakagami distribution has a simple dependence on r, and thus is often used in tractable analysis of fading performance [6]. When the envelope r is assumed to be Nakagami-distributed, the squared-envelope r^2 has a Gamma distribution [4]. The Nakagami distribution is capable of modeling more severe fading than Rayleigh fading by selecting ½ < m < 1. However, due to the lack of physical basis, the Nakagami distribution is not as popular as the Rayleigh and Ricean fading models in mobile communications.

1.5. Suzuki Fading:

The Suzuki model [7] is a statistical model that gives the composite distribution due to log-normal shadowing and Rayleigh fading. This model is particularly useful for link performance evaluation of slow moving or stationary MSs, since the receiver has difficulty in averaging the effects of fading. It is widely accepted for the signal envelope received in macro cellular mobile channels with no LOS path.

1.6. Doppler Fading:

Multipath components lead to delay dispersion, while the Doppler effect leads to frequency dispersion for a multipath propagation. Doppler spread is also known as time-selective spread. Frequency-dispersive channels are known as time-selective fading channels. Signals are distorted in both the cases. Delay dispersion is dominant at high data rates, while frequency dispersion is dominant at low data rates. The two dispersions are equivalent, since the Fourier transform can be applied to move from the time domain to the frequency domain. These distortions cannot be eliminated by just increasing the transmit power, but can be reduced or eliminated by equalization or diversity.

II. A Review On The Wide Sense Stationary Uncorrelated Scattering Model

Wireless channels are time-variant, with an impulse response h (t,τ), and can be modelled by using the theory of linear time-variant systems. Because most wireless channels are slowly time-varying, or quasi-static, this enables the use of many concepts of linear time-invariant (LTI) systems. By performing the Fourier transform to the absolute time t, or the delay τ, or both, we obtain the delay Doppler-spread function S (v,τ), the time-variant transfer function H(t,f), or the Doppler-variant transfer function B(v,f), respectively. The stochastic model of wireless channels is a joint pdf of the complex amplitudes for any τ and t. The ACF is usually used to characterize the complex channel. The ACF of the received signal, y(t)=x(t) * h(t,τ), for a linear time-variant system is given by

$$R_{yy}(t,t') = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{xx}(t-\tau, t'-\tau')R_h(t,t',\tau,\tau')d\tau d\tau' \quad (11)$$

where the ACFs are

$$R_{XX}(t-\tau, t'-\tau') = E[x^*(t-\tau)x(t'-\tau')] \quad (12)$$

$$R_h(t,t',\tau,\tau') = E[h^*(t,\tau)h(t',\tau')] \quad (13)$$

A stochastic process is said to be strictly stationary if all its statistical properties are independent of time. When only the mean is independent of time while the autocorrelation depends on the time difference $\Delta t = t' - t$, such a process is said to be a wide sense stationary process. The popular WSSUS (wide sense stationary, uncorrelated

scattering) model is based on the dual assumption: wide sense stationarity and uncorrelated scatters. The assumptions of wide sense stationarity states that the ACF of the channel is determined by the difference Δt, that is,

$$R_h(t,t',\tau,\tau') = R_h(\Delta t,\tau,\tau') \tag{14}$$

Thus, the statistical properties of the channel do not change over time, and the signals arriving with different Doppler shifts are uncorrelated

$$R_h(v,v',\tau,\tau') = P_s(v,\tau,\tau')\delta(v-v') \tag{15}$$

where P_s the scattering function, giving the Doppler power spectrum for a multipath channel with different path delays τ.

2.1. Delay Spread:

The delay power spectral density or power delay profile (PDP), $P_h(\tau)$, is obtained by integrating the scattering function $P_s(\tau,v)$ over the Doppler shift v. The PDP can be calculated by the following expression

$$P_h(\tau) = \sum_{n=1}^{N} P_n \delta(\tau-\tau_n) = \int_{-\infty}^{\infty} \left| h(t,\tau) \right|^2 dt \tag{16}$$

where $P_n = a_n^2$, a_n being the amplitude of the nth delay, and the second equality holds if ergodicity holds. First arrival delay τ_A is the delay of the first arriving signal, all the other delays are known as excess delays, and the maximum excess delay τ_{max} is the delay corresponding to a specified power threshold. Delay spread leads to frequency-selective fading, as the channel function resembles a tapped-delay filter. A general rule of thumb is that $\tau_{max} \approx 5\sigma_\tau$. The PDP has been modelled in order to understand the channel behaviour and to evaluate the performance of equalizers. There are many measurements of indoor and outdoor channels [8]. The one-sided exponential profile is a suitable model for both indoor and urban channels

$$P_h(\tau) = \frac{P_T}{\sigma_\tau} e^{-\frac{\tau}{\sigma_\tau}}, \tau \geq 0 \tag{17}$$

where P_T is the total received power. When the excess delay spread exceeds the symbol duration by 10% to 20%, an equalizer may be required. The average delay and the delay spread of a channel diminish with decreasing cell size due to shorter propagation path.

2.2 Correlation Coefficient:

The correlation coefficient of two signals is usually defined with respect to the signal envelopes x and y

$$\rho = \rho_{xy} = \frac{E[xy]-E[x]E[y]}{\sqrt{(E[x^2]-E[x]^2)(E[y^2]-E[y]^2)}} \tag{18}$$

For two statistically independent signals, ρ=0; when ρ is below a threshold such as 0.5, the signals are typically considered effectively as decorrelated. For a channel with PDP of type (16), assuming a classical Doppler spectrum for all the components, the correlation coefficient of two signals with a temporal separation Δt and a frequency separation Δf is given by [9, 2]

$$\rho_{xy}(\Delta t,\Delta f) = \frac{J_0^2(2\pi v_{max}\Delta t)}{1+(2\pi\sigma_\tau\Delta f)^2} \tag{19}$$

where $J_0(x)$ is the zero-order Bessel function of the first kind and v_{max} is the maximum Doppler frequency.

2.3. Channel Coherent Bandwidth:

The channel coherence bandwidth B_c is defined as the maximum frequency difference $(\Delta f)_{max}$ that limits the correlation coefficient ρ to be smaller than a given threshold, typically 0.7 is given as follows

$$B_c = (\Delta f)_{max} = \frac{1}{2\pi\sigma_\tau} \tag{20}$$

Due to the uncertainty relation between the Fourier transform pairs, there is an uncertainity relation between B_c and the rms relay spread σ_τ [2] as follows

$$B_c \geq \frac{1}{2\pi\sigma_\tau} \tag{21}$$

That is, both B_c and σ_τ can be used to characterize the channel, and they are in inverse proportion; although usually they can be related by the approximation $B_c \approx \frac{1}{\sigma_\tau}$, they cannot be exactly derived from each other.

2.4. Doppler Spread And Channel Coherent Time:

The Doppler power spectral density $P_B(v)$ is obtained by integrating the scattering function $P_s(\tau,v)$ over the time delay τ. Analogous to the derivation of the average channel delay $\bar{\tau}$ and rms delay spread σ_τ can be derived as the first- and second-order moments of $P_B(v)$. The Doppler spread corresponds to time-selective fading. The channel coherence time T_c is also defined according to (19). The coherent time measures how fast the channel changes in time. A large coherent time corresponds to a slow channel fluctuation. The coherence time is defined in a manner similar to that of the coherent bandwidth. It is defined as the time delay for which the signal autocorrelation coefficient reduces to 0.7.

2.5. Angle Spread And Coherent Distance:

The channel model can also include the directional information such as the DoA and DoD of the multipath components into its impulse response, leading to the double-directional impulse response. Analogous to the non directional case, a number of power spectrums such as the double directional delay power spectrum, angular delay power spectrum, angular power spectrum, and azimuthal spread can be defined [10]. Such a directional channel model is especially useful for multiple antenna systems. Angle spread at the receiver is the spread in DoA of the multipath components at the receive antenna array, while angle spread at the transmitter is the spread in DoDs of the multipath components that finally arrive at the receiver. Denoting the DoA by Θ, the angle power spectrum or power angular profile $P_A(\Theta)$ is given by the following expression

$$P_A(\theta) = \sum_{n=1}^{N} P_n \delta(\theta - \theta_n) \tag{22}$$

III. Non-Linear Modulation Methods With Memory-Cpfsk And Cpm

A class of digital modulation methods in which the phase of the signal is constrained to be continuous. This constraint results in a phase or frequency modulator that has memory. The modulation method is also non-linear.

3.1. CONTINUOUS-PHASE FSK (CPFSK): An conventional FSK signal is generated by shifting the carrier by an amount $f_n = \frac{1}{2}\Delta f I_{n,} I_n = \pm 1, \pm 3, ..., \pm(M-1),$ to reflect the digital information that is being transmitted. The switching from one frequency to another may be accomplished by having $M=2^k$, separate oscillators tuned to the desired frequencies and selecting one of the M frequencies according to the particular k-bit symbol that is to be transmitted in a signal interval of duration T=k/R seconds. However, such abrupt switching from one oscillator output to another in successive signaling intervals results in relatively large spectral side lobes outside the main spectral band of the signal and, consequently, this method requires a large frequency band for transmission of the signal.

To avoid the use of signals having large spectral side lobes, the information-bearing signal frequency modulates a single carrier whose frequency is changed continuously. The resulting frequency-modulated signal is phase-continuous and, hence, it is called continuous-phase FSK (CPFSK)[11]. This type of FSK signal has memory because the phase of the carrier is constrained to be continuous. In order to represent a CPFSK signal, we begin with a PAM signal and it is expressed as follows

$$d(t) = \sum_n I_n g(t - nT) \tag{23}$$

where $\{I_n\}$ denotes the sequence of amplitudes obtained by mapping k-bits blocks of binary digits from the information sequences $\{a_n\}$ into the amplitudes levels $\pm 1, \pm 3, ..., \pm(M-1)$ and g(t) is a rectangular pulse of x amplitude 1/2T and duration T seconds. The signal d(t) is used to frequency-modulate the carrier. Consequently, the equivalent low-pass waveform v(t) is expressed as follows

$$v(t) = \sqrt{\frac{2\varepsilon}{T}} \exp\{ j\left[4\pi T f_d \int_{-\infty}^{t} d(\tau) d\tau + \phi_0 \right]\} \tag{24}$$

where f_d is the peak frequency deviation and ϕ_0 is the initial phase of the carrier. The carrier-modulated signal corresponding to the (24) may be expressed as the following

$$s(t) = \sqrt{\frac{2\varepsilon}{T}} \cos[2\pi f_c t + \phi(t;I) + \phi_0] \tag{25}$$

$$\text{where } \phi(t;I) = 4\pi T f_d \int_{-\infty}^{t} d(\tau) d\tau$$

$$\phi(t;I) = 4\pi T f_d \int_{-\infty}^{t} [\sum_n I_n g(\tau - nT)] d\tau \tag{26}$$

It is noted that, although d(t) is contains discontinuities, the integral of d(t) is continuous. Hence, we have a continuous-phase signal. The phase of the carrier in the interval $nT \le t \le (n+1)T$ is determined by integrating (26). Thus the following equations are obtained

$$\phi(t;I) = 2\pi f_d T \sum_{k=-\infty}^{n-1} I_k + 2\pi f_d (t - nT) I_n \tag{27}$$

$$= \theta_n + 2\pi h I_n q(t - nT) \tag{28}$$

Where h, Θ_n, and q(t) are defined as

$$h = 2 f_d T \tag{29}$$

$$\theta_n = \pi h \sum_{k=-\infty}^{n-1} I_k \tag{30}$$

$$q(t) = \begin{cases} 0 \; for (t < 0) \\ t / 2T for (0 \le t \le T) \\ 1/2 \; for (t > T) \end{cases} \tag{31}$$

The observation is that Θ_n represents the accumulation (memory) of all symbols up to time (n-1) T. The parameter h is called the modulation index.

3.2. CONTINUOUS-PHASE MODULATION (CPM):

When expressed in the form of (28), CPFSK becomes a special case of a general class of continuous-phase modulated (CPM) signals in which the carrier phase is written as follows:

$$\phi(t;I) = 2\pi \sum_{k=-\infty}^{n} I_k h_k q(t - kT), nT \le t \le (n+1)T \tag{32}$$

Where $\{I_k\}$ is the sequence of M-ary information symbols selected from the alphabet $\pm 1, \pm 3, ... \pm (M-1)$, $\{h_k\}$ is a sequence of modulation indices, and q(t) is some normalized waveform shape. When h_k=h for all k, the modulation index is fixed for all symbols. When the modulation index varies from one symbol to another, the CPM signal is multi-h. In such a case, the $\{h_k\}$ are made to vary in a cyclic manner through a set of indices. The waveform q(t) may be represented in general as the integral of some pulse g(t), i.e.,

$$q(t) = \int_0^t g(\tau) d\tau \tag{33}$$

If g(t)=0 for t>T, the CPM signal is called full response CPM. If $g(t) \neq 0$ for t>T, the modulated signal is called partial response CPM. It is apparent that an infinity variety of CPM signals can be generated by choosing different pulse shapes g(t) and by varying the modulation index h and the alphabet size M. It is noted that the CPM signal has memory that is introduced through the phase continuity [12]. For L>1, additional memory is introduced in the CPM signal by the pulse g(t) and by varying the modulation index h and the alphabet size M.

IV. Performance Analysis Of Rician Fading Channels Using Cpfsk And Cpm In Simulink:

The environment is created as shown in the fig. 1.and fig.2. respectively using Simulink tool.

5.1 RANDOM INTEGER GENERATOR: The random integer generator generates random uniformly distributed integers in the range [0, M-1], where M is the M-ary number.

5.2. INTEGER TO BIT CONVERTER: In the integer to bit convertor unit, a vector of integer-valued or fixed valued type is mapped to a vector of bits. The number of bits per integer parameter value present in the integer to bit convertor block defines how many bits are mapped for each integer-valued input. For fixed-point inputs, the stored integer value is used. This block is single-rated and so the input can be either a scalar or a frame-based column vector. For sample-based scalar input, the output is a 1-D signal with 'Number if bits per integer' elements. For frame-based column vector input, the output is a column vector with length equal to 'Number of bits per integer' times larger than the input signal length.

5.3 DIFFERENTIAL ENCODER: Differential encoder differentially encodes the input data. The differential encoder object encodes the binary input signal within a channel. The output is the logical difference between the current input element and the previous output element.

5.4 CONVOLUTIONAL INTERLEAVER: This block permutes the symbols in the input signal. Internally, it uses a set of shift registers. The delay value of the kth shift register is (k-1) times the register length step parameter. The number of shift registers is the value of the rows of shift registers parameter.

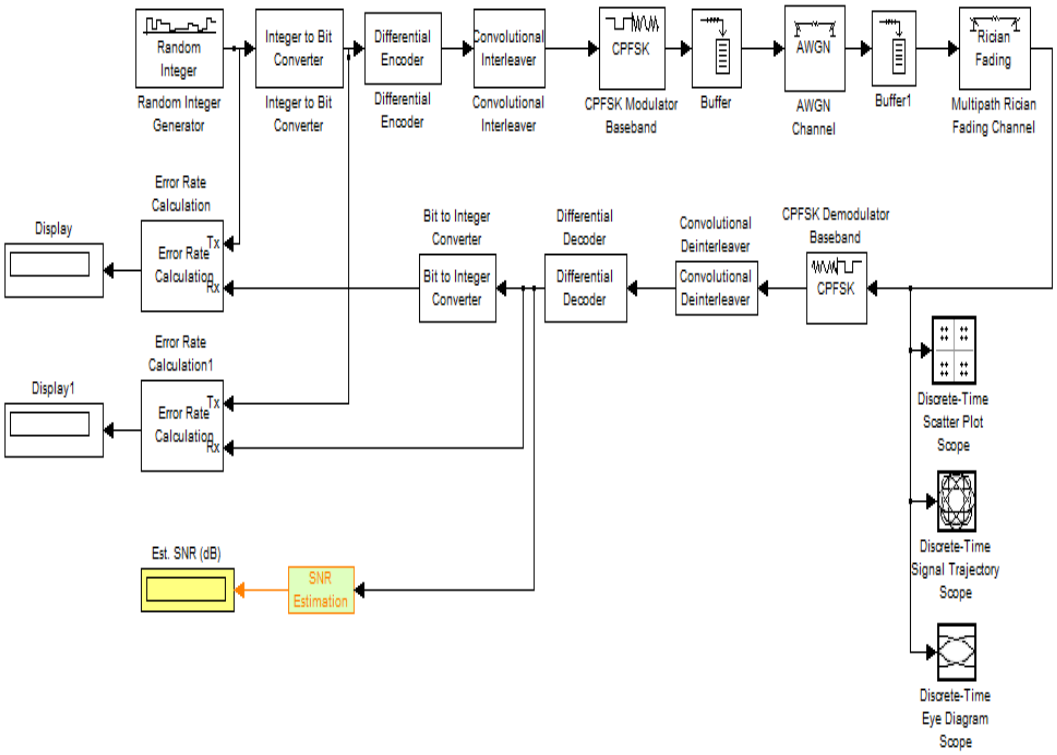


Fig. 1. Screenshot for the performance analysis of Rician fading channels using CPFSK modulation in Simulink

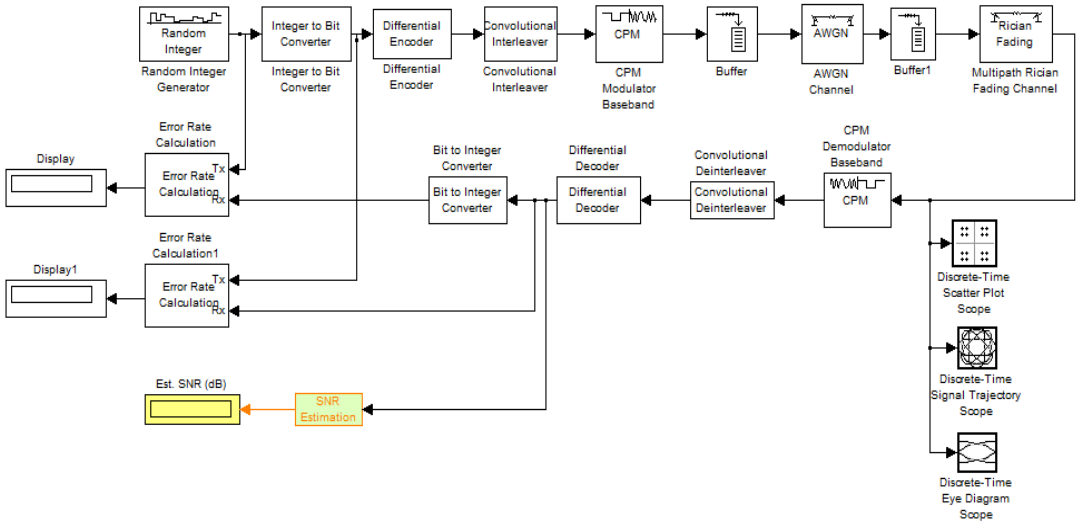


Fig. 2. Screenshot for the performance analysis of Rician fading channels using CPM modulation in Simulink

5.5 CPFSK MODULATOR BASEBAND: This block modulates the input signal using the continuous phase frequency shift keying method. The input can be either bits or integers.
5.6.CPFSK DEMODULATOR BASEBAND: This block demodulates the CPFSK modulated input signal using the Viterbi algorithm.
5.7 CPM MODULATOR BASEBAND: This block gives the output of the complex envelope representation of the selected continous phase modulation. The input can be either bits or integers.
5.8 CPM DEMODULATOR BASEBAND: This block demodulates the CPM modulated input signal using the Viterbi algorithm.
5.7 BUFFER: The buffer converts scalar samples to a frame output at a lower sample rate. The conversion of a frame to a larger size or smaller size with optional overlap is possible. It is then passed to the multipath Rician fading channel.
5.8 CONVOLUTIONAL DEINTERLEAVER: The Convolutional deinterleaver block recovers a signal that was interleaved using the Convolutional interleaver block.
5.9 DIFFERENTIAL DECODER: The differential decoder block decodes the binary input signal.
5.10 BIT TO INTEGER CONVERTER: The bit to integer converter maps a vector of bits to a corresponding vector of integer values. The number of bits per integer parameter defines how many bits are mapped for each output.
5.11 ERROR RATE CALCULATION: The error rate calculation is done by computing the error rate of the received data by comparing it to a delayed version of the transmitted data.
5.12 SIGNAL TRAJECTORY SCOPE: The discrete-time signal trajectory scope is used to display a modulated signal constellation in its signal space by plotting the in phase component versus the quadrature component.
5.13 SCATTER PLOT SCOPE: The discrete-time scatter plot scope is used to display a modulated signal constellation in its signal space by plotting the in phase component versus the quadrature component.
5.14 EYE DIAGRAM SCOPE: The discrete-time eye diagram scope displays multiple traces of a modulated signal to reveal the modulation characteristics such as pulse shaping, as well as channel distortions of the signal.
5.15 SNR ESTIMATION: The SNR estimation block gives the estimated SNR in decibels.
5.16 DISPLAY: This unit gives the total number of bits transmitted, the number of errors and finally displays the Bit Error Rate.

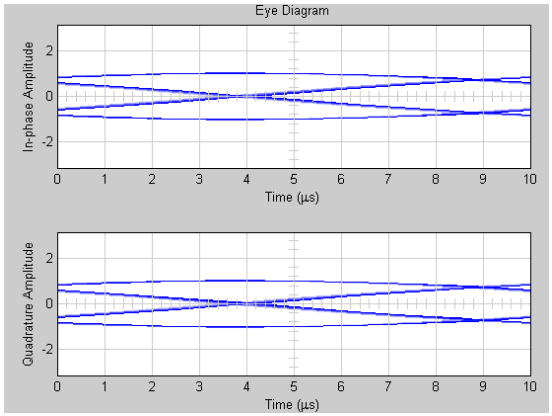


Fig. 1.Eye diagram for the performance analysis of Rician Fading Channels in CPFSK scheme

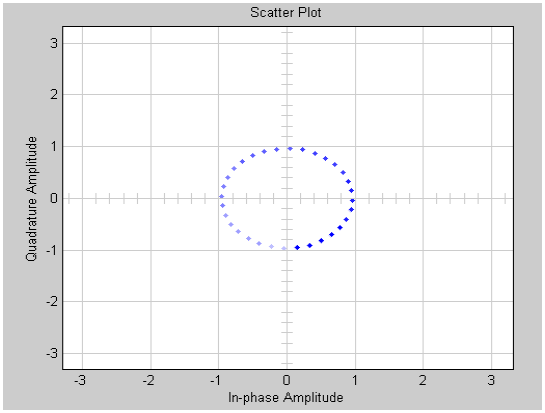


Fig. 2. Scatter plot for the performance analysis of Rician Fading Channels in CPFSK scheme

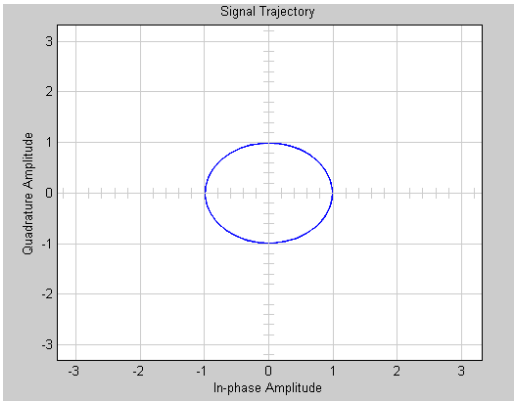


Fig. 3. Signal Trajectory for the performance analysis of Rician Fading Channels in CPFSK scheme

Table.1. BER for Rician fading channels in CPFSK modulation scheme

K_r(Ricean factor)	SNR(dB)	BER
100	10	0.009079
500	50	0.009048
1000	100	0.008895

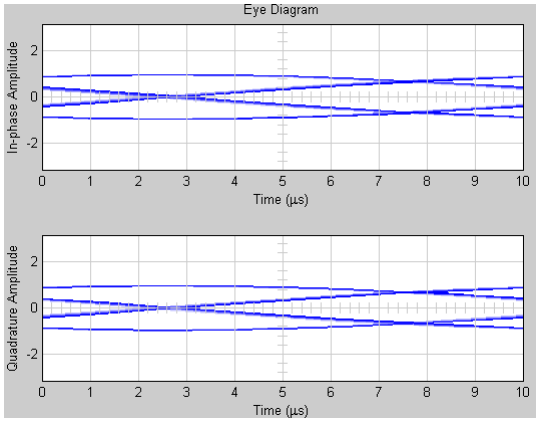


Fig. 4. Eye diagram for the performance analysis of Rician Fading Channels in CPM scheme

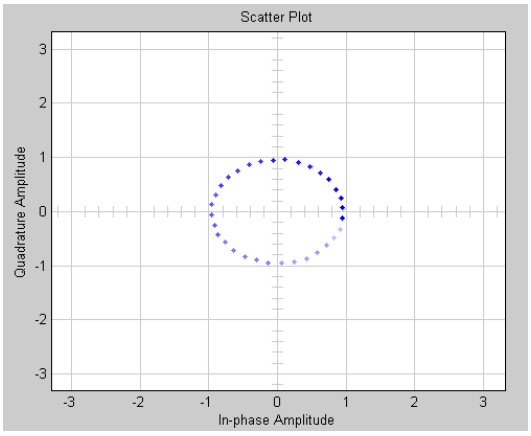


Fig. 5. Scatter plot for the performance analysis of Rician Fading Channels in CPM scheme

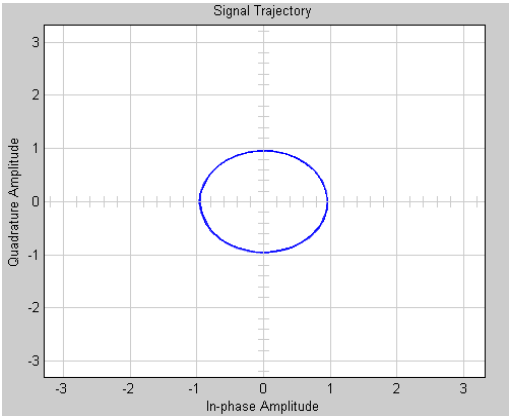


Fig. 6. Signal Trajectory for the performance analysis of Rician Fading Channels in CPM scheme

Table.2. BER for Rician fading channels in CPM schemes

K_r(Ricean factor)	SNR(dB)	BER
100	10	0.009317
500	50	0.009288
1000	100	0.009126

V. Conclusion

A short survey on different types of channel fading is provided followed by a review on wide sense stationary uncorrelated scattering model. The analysis of Rician fading channels in CPFSK modulation schemes and CPM modulation schemes is discussed and the results are provided. It is evident from table 1 and table 2 that when the Ricean factor (K_r) and the Signal-to-Noise Ratio is increased then the resulting Bit Error Rate gradually decreases. For high values of K_r, a very low bit error rate is achieved in CPFSK modulation scheme than in the CPM modulation scheme. The eye diagram, signal trajectory diagram and the scatter plot diagrams have also been provided for the scenario. Future works may include finding the bit error rate by evaluating the performance of Rician fading channels by using various modulation schemes and by altering the Ricean factor parameters.

References:

[1] A.Goldsmith, *Wireless Communication* (Cambridge, UK:Cambridge University Press,2005)
[2] A.F.Molisch, *Wireless Communications* (Chichester: Wiley-IEEE,2005).
[3] S.O.Rice, Statistical properties of a sine wave plus random noise. *Bell Syst.Tech.J.*, 27:1 (1948), 109-157.
[4] G.L.Stuber, *Principles of Mobile Communication*, 2nd edn (Boston, MA:Kluwer, 2001).
[5] M.Nakagami, The m-distribution: a general formula of intensity distribution of rapid fading. In W.C.Hoffman, ed., *Statistical Methods in Radio Wave Propagation* (Oxford:Pergamon Press, 1960), pp.3-36.
[6] M.K.Simon and M.-S.Alouini, *Digital Communications over Fading Channels*, 2nd edn (New York: Wiley, 2005).
[7] H.Suzuki, A statistical model for urban radio propagation, *IEEE Trans. Commun*., 25:7 (1977), 673-680.
[8] J.Chuang, The effects of time delay spread on portable radio communications channels with digital modulation. *IEEE J.Sel.Areas.Commun*., 5:5(1987), 879-889.
[9] W.C.Jakes, Jr. (ed.), *Microwave Mobile Communications* (New York: Wiley, 1974).
[10] B.H.Fleury, First- and second-order characterization of direction dispersion and space selectivity in the radio channel. *IEEE Trans. Inf.Theory*, 46:6 (2000), 2027-2044.
[11] J.M.Wozencraft and I.M.Jacobs, *Principles of Communication Engineering*, Wiley, New York 1965.
[12] C.R.Cahn, "Combined Digital Phase and Amplitude Modulation Communication Systems," *IRE Transactions on Communications Systems*, Vol. CS-8, pp.150-154, 1960.

IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 11, Issue 4 (May. - Jun. 2013), PP 37-43
www.iosrjournals.org

Implementation of Matching Tree Technique for Online Record Linkage

B.Rakesh[1], M.Madhavi[2] ,CH.Venkata Ratnam[3], C.Ravindra Murthy[4]
1Asst.professor, Sreevidyanikethan engg.college, Tirupathi, India
2 Asst.professor, ASRA, Hyderabad, India,
3 Asst.professor, Holy Mery institute of engg&technology, hyd,India
4 Asst.professor, Sreevidyanikethan engg.college, Tirupathi, India,

Abstract: *The task of finding records refering to the same entity across heterogeneous data sources is known as record linkage .the necessity to consolidate the information located in different data sources has been widely documented in recent years. For the purpose of completing this goal, several types of problems ,must be solved by an organization .when same real world substance is used by different identifiers in different sources entity heterogeneity problem will arise. For solving the entity heterogeneity problem statistical record linkage technique could be used. However, the use of such techniques for online record linkage could pose a tremendous communication bottleneck in a distributed environment (where entity heterogeneity problem often encountered). In order to resolve this issue, we develop a matching tree, similar to a decision tree, and use it to propose techniques that reduce the communication overhead significantly, while providing matching decisions that are guaranteed to be the same as those obtained using the conventional linkage technique. These techniques have been implemented, and experiments with real-world and synthetic databases show significant reduction in communication overhead.*

I. Introduction

The record-linkage problem identifying and linking duplicate records arises in the context of data cleansing, which is a necessary pre-step to many database applications. Databases frequently contain approximately duplicate fields and records that refer to the same real-world entity, but are not identical.

Importance of data linkage in a variety of data-analysis applications, developing effective an efficient techniques for record linkage has emerged as an important problem. It is further evidenced by the emergence of numerous organizations (e.g., Trillium, First Logic , Vality, Data Flux) that are developing specialized domain specific record-linkage and data-cleansing tools.

The data needed to support these decisions are often scattered in heterogeneous distributed databases. In such cases, it may be necessary to link records in multiple databases so that one can consolidate and use the data pertaining to the same real world entity. If the databases use the same set of design standards, this linking can easily be done using the primary key (or other common candidate keys). However, since these heterogeneous databases are usually designed and managed by different organizations (or different units within the same organization), there may be no common candidate key for linking the records. Although it may be possible to use common non key attributes (such as name, address, and date of birth) for this purpose, the result obtained using these attributes may not always be accurate. This is because non key attribute values may not match even when the records represent the same entity instance in reality.

The databases exhibiting entity heterogeneity are distributed, and it is not possible to create and maintain a central data repository or warehouse where pre computed linkage results can be stored. A centralized solution may be impractical for several reasons. First, if the databases span several organizations, the owner ship and cost allocation issues associated with the warehouse could be quite difficult to address. Second ,even if the warehouse could be developed, it would be difficult to keep it up-to-date. As updates occur at the operational databases, the linkage results would become stale if they are not updated immediately .This staleness maybe unacceptable in many situations .For instance, in a criminal investigation, one maybe interested in the profile of crimes committed in the last 24 hours within a certain radius of the crime scene. In order to keep the warehouse current, the sites must agree to transmit incremental changes to the data warehouse on a real-time basis. Even if such an agreement is reached, it would be difficult to monitor and enforce it. For example, a site would often have no incentive to report the insertion of a new record immediately. Therefore, these changes are likely to be reported to the warehouse at a later time, thereby increasing the staleness of the linkage tables and limiting their usefulness. In addition, the overall data management tasks could be prohibitively time-consuming, especially in situations where there are many databases, each with many records, undergoing real-time changes. This is because the warehouse must maintain a linkage table for each pair of sites, and must update them every time one of the associated databases changes.

The participating sites allow controlled sharing of portions of their databases using standard database queries, but they do not allow the processing of scripts, stored procedures, or other application programs from another organization .The issue here is clearly not one of current technological abilities, but that of management and control. If the management of an organization wants to open its databases to outside scripts from other organizations, there are, of course, a variety of ways to actually implement it. However, the decision to allow only a limited set of database queries (and nothing more) is not based on technological limitations; rather it is often a management decision arising out of security concerns. More investment in technology or a more sophisticated scripting technique, therefore, is not likely to change this situation. A direct consequence of this fact is that the local site cannot simply send the lone enquiry record to the remote site and ask the remote site to perform the record linkage and send the results back

1.1 OBJECTIVE

If the databases use the same set of design standards, this linking can easily be done using the primary key (or other common candidate keys). However, since these heterogeneous databases are usually designed and managed by different organizations (or different units within the same organization), there may be no common candidate key for linking the records. Although it may be possible to use common non key attributes (such as name, address, and date of birth) for this purpose, the result obtained using these attributes may not always be accurate. This is because non key attribute values may not match even when the records represent the same entity instance in reality.

II. Literature Survey:

2.1 Linked Record Health Data Systems:

The goal of record linkage is to link quickly and accurately records that correspond to the same person or entity. Whereas certain patterns of agreements and disagreements on variables are more likely among records pertaining to a single person than among records for different people, the observed patterns for pairs of records can be viewed as arising from a mixture of matches and non matches. Mixture model estimates can be used to partition record pairs into two or more groups that can be labeled as probable matches (links) and probable non matches. A method is proposed and illustrated that uses marginal information in the database to select mixture models, identifies sets of records for clerks to review based on the models and marginal information, incorporates clerically reviewed data, as they become available, into estimates of model parameters, and classifies pairs as links, non links, or in need of further clerical review. The procedure is illustrated with five datasets from the U.S. Bureau of the Census. It appears to be robust to variations in record-linkage sites. The clerical review corrects classifications of some pairs directly and leads to changes in classification of others through re estimation of mixture models

2.2 Efficient Private Record Linkage:

Record linkage is the computation of the associations among records of multiple databases. It arises in contexts like the integration of such databases, online interactions and negotiations, and many others. The autonomous entities who wish to carry out the record matching computation are often reluctant to fully share their data. In such a framework where the entities are unwilling to share data with each other, the problem of carrying out the linkage computation without full data exchange has been called private record linkage. Previous private record linkage techniques have made use of a third party. We provide efficient techniques for private record linkage that improve on previous work in that (i) they make no use of a third party; (ii) they achieve much better performance than that of previous schemes in terms of execution time and quality of output (i.e., practically without false negatives and minimal false positives). Our software implementation provides experimental validation of our approach and the above claims.

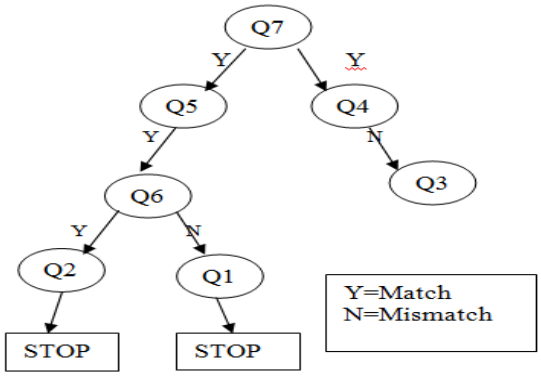


FIG 1: A sample tree showing the attribute aquistion order

The RTS technique used in that framework is scenario-slicing which is not safe. Additionally, the mechanism with which they connect services and applications together is performed manually through the creation and use of scenario-slices. Our approach, in contrast, is safe, works with Java Web services without requiring any additional information other than what is provided by Axis and the source code, and is This framework is described as rapid and adaptive end-to-end regression testing. provided programmatically.

The code-based RTS for component-based software by Harrold's team is also a significantly related work. Their approach revolves around what the authors refer to as met content. This meta-content is additional data expected to be provided by vendors in lieu of source code in such cases where that content may be restricted by patent or copyright law. Their required meta-content is threefold: edge coverage achieved by the test suite with respect to the component, component version, and a way to query the component for the edges affected by the changes between the two given versions. This code-based approach is a white-box testing using code where applicable and using meta-content where necessary. This is a safe RTS. However, it cannot be directly applied to Web services due to Web services being composable in nature. In a Web services world, an application could call a service, which in turn calls other services. In such an environment, an application can be affected not only by the changes in the services it directly calls on but also by the changes in the services that are used indirectly. Thus, just querying about changes in every called component (service) is not sufficient.

TRANSFERED	**CONCURRENT**	**SEQUENTIAL**
Identifier	Not applicable	Sequential identifier Acquistion
Attribute	Concurrent attribute acquistion	Sequential attribute Acquistion

Fig 2: Possible tree based linkage techniques

2.3 A Survey of Approaches to Automatic Schema Matching:

Schema matching is a basic problem in many database application domains, such as data integration, Ebusiness, data warehousing, and semantic query processing .In current implementations, schema matching is typically performed manually, which has significant limitations. On the other hand, previous research papers have proposed many techniques to achieve a partial automation of the match operation for specific application domains. We present a taxonomy that covers many of these existing approaches, and we describe the approaches in some detail. In particular ,we distinguish between schema-level and instance-level, element-level and structure-level, and language-based and constraint-based matchers. Based on our classification we review some previous match implementations thereby indicating which part of the solution space they cover .We intend our taxonomy and review of past work to be useful when comparing different approaches to schema matching, when developing a new match algorithm, and when implementing a schema matching component.

One empirical approach is to perform a case study on an existing program that has several versions and existing test suites. The changes to such a program would be "real". However, under this approach, the circumstances under which evolution and testing occur are not controlled. For example, the changes may have different sizes and implications, and there may be only one test suite per version. Hence, it could be difficult to draw valid inferences about causality or about the ability of results to generalize. A second empirical approach is to perform a controlled experiment in which changes are simulated. The advantage of such an experiment is that the independent variables of interest (number of changes, location of changes, and test suite) can be manipulated to determine their impact on dependent variables . This lets us apply different values to the independent variables in a controlled fashion, so that results are not likely to depend on unknown or uncontrolled factors. The primary weakness of this approach, however, is the threat to external validity posed by the change simulation. Because each approach has different advantages and disadvantages, we employed both. 3 Study 1 We present our controlled experiment first.

A Comparison of Fast Blocking Methods for Record Linkage

The task of linking databases is an important step in an increasing number of data mining projects, because linked data can contain information that is not available otherwise, or that would require time-consuming and expensive collection of specific data. The aim of linking is to match and aggregate all records that refer to the same entity. One of the major challenges when linking large databases is the efficient and accurate classification of record pairs into matches and non-matches. While traditionally classification was based on manually-set thresholds or on statistical procedures, many of the more recently developed classification methods are based on supervised learning techniques. They therefore require training data, which is often not available in real world situations or has to be prepared manually, an expensive, cumbersome and time-consuming process.

We generated 30 versions for each change level, each with a randomly defined change probability. One of our research questions concerns differences in types of code coverage information. For this experiment, we selected two types of coverage: statement coverage, which tracks the statements executed by each test, and function coverage, which tracks the functions executed by each test. In both cases, we track not frequency of execution, but just whether the component was or was not executed. To measure coverage, we used the Aristotle analysis system. Some branch modifications caused Space to crash. One option for handling this was to discard tests that caused crashes; however, this could discard valuable coverage information, causing us to underestimate the effects of changes

Moreover, in practice, such tests would not be discarded. A second option was to modify Space to capture termination signals and record coverage data before failing. We chose this second option. Similarly, we inserted traps into Space to halt execution on modifications that lead to infinite loops. (In practice, such loops would be terminated by human intervention, our traps simulate this.) 3.2 Experiment Design To put Space and its modifications through all desired experimental conditions we selected a factorial design, considering all combinations of versions, test suites, and change levels. In more formal terms, our design constitutes a completely randomized factorial design with three treatments: (T1) versions, with 30 levels (excluding the baseline), (T2) test suites, with 50 levels, and (T3) change level, with five levels. Given this design, we generated 7500 observations (excluding the 50 observations for the baseline) at the statement level and 7500 observations at the function level.3 Each observation contained the four metrics presented in Section 2.1, for each level of the three factors. 3.3 Results and Analysis First, we examine the data graphically. Box plots for each metric at the baseline and the five change levels are presented in Figure 1.4 3 The column of graphs on the left depicts results for statement coverage; the column on the right depicts results for function coverage; the four graphs in each column depict results for each of the four metrics (MD, CC, CAC, CAT), respectively.

The x-axes represents the baseline and five change levels. The y-axes represent the values measured for each metric. Each individual box plot (other than the plots for baselines) indicates the distribution of the 1500 observations (50 test suites times 30 versions) for one change level, for a given metric and coverage type. (Plots for baselines, where no change levels are involved, represent 50 observations.) As the plots indicate, MD (matrix density) decreased as change level increased. Both statement and function coverage presented the same trend. However, MD based on function coverage information exhibited a greater rate of decrease and greater deviation (as indicated by the size of the box plots). CC (component coverage) also decreased as change level increased. As expected, CC at the function level was higher at lower change levels, but at higher change levels it rapidly decreased to values similar to CC at the statement level. (Recall from Section 3.1 that Space contains some unreachable code and functions; this explains why CC is not 100% for either type of coverage, even for the baseline version.) It is interesting to note that the mean CC decreases more than 10% if even 1% of the branches are affected by a change. That decreased percentage rose to 20% when 2% of the branches were affected by a change.

The author has previously presented a novel two-step approach to automatic record pair classification. In the first step of this approach, training examples of high quality are automatically selected from the compared record pairs, and used in the second step to train a support vector machine (SVM) classifier. Initial experiments showed the feasibility of the approach, achieving results that outperformed *k*-means clustering. In this paper, two variations of this approach are presented. The first is based on a nearest neighbor classifier, while the second improves a SVM classifier by iteratively adding more examples into the training sets. Experimental results show that this two-step approach can achieve better classification results than other unsupervised approaches.

A Method for Calibrating False-Match Rates in Record Linkage

Specifying a record-linkage procedure requires both (1) a method for measuring closeness of agreement between records, typically a scalar weight, and (2) a rule for deciding when to classify records as matches or non matches based on the weights. Here we outline a general strategy for the second problem, that is, for accurately estimating false-match rates for each possible cutoff weight. The strategy uses a model where the distribution of observed weights is viewed as a mixture of weights for true matches and weights for false matches. An EM algorithm for fitting mixtures of transformed-normal distributions is used to find posterior modes; associated posterior variability is due to uncertainty about specific normalizing transformations as well as uncertainty in the parameters of the mixture model, the latter being calculated using the SEM algorithm. This mixture-model calibration method is shown to perform well in an applied setting with census data. Further, a simulation experiment reveals that, across a wide variety of settings not satisfying the model's assumptions, the procedure is slightly conservative on average in the sense of overstating false-match rates, and the one-sided confidence coverage (i.e., the proportion of times that these interval estimates cover or overstate the actual false-match rate) is very close to the nominal rate.

Applying Model Management to Classical Meta Data Problems

We wish to measure the evidence that a pair of records relates to the same, rather than different, individuals. The paper emphasizes statistical models which can be fitted to a file of record pairs known to be correctly matched, and then used to estimate likelihood ratios. A number of models are developed and applied to UK immigration statistics. The combination of likelihood ratios for possibly correlated record fields .

To determine which change levels actually differed, we performed a Bonferroni analysis (see Table 5). For each metric, the table presents the mean, and a grouping letter to represent the Bonferroni results (change levels with the same letter are not significantly different). Although the averages for all metrics seem to grow closer as change level increases, we found that only in one case (comparing CAT at the 10% and 15% levels for statement coverage) are the metrics not significantly different. Finally, we performed an analysis to compare the metrics means between types of coverage information (see Table 6). The ANOVA indicated that, when compared at each change level, statement and function coverage information generated significantly different metrics. As expected, most of the comparisons indicated that function coverage information tends to have higher values across all change levels for MD and CC. This is intuitively plausible when we consider that it is common for a change that alters statement coverage not to modify the functional control flow.

The interpretation of the change metrics was not as intuitive. For CAC, the analysis indicated that function coverage information was more susceptible to changes in the program than was statement level information. That means that although CC may have remained similar between versions, the functions that were covered were different. CAT on the other hand shows different results depending on the change level. At the 1% and 2% levels, function coverage information follows the same trend as CAC. This trend, however, is reversed at higher change levels, possibly reflecting different thresholds on maximum change.

Record Linkage Current Practice and Future Directions

Record linkage is the task quickly and accurately identifying records corresponding to the same entity from one or more data sources. Record linkage is also known as data cleaning, entity reconciliation or identification and the merge/purge problem. This paper presents the "standard" probabilistic record linkage model and the associated algorithm. Recent work in information retrieval, federated database systems and data mining have proposed alternatives to key components of the standard algorithm. The impact of these alternatives on the standard approach are assessed.

III. Results:

Fig:User login page

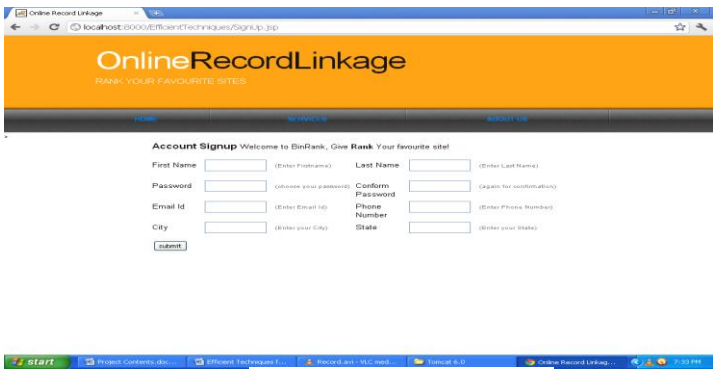
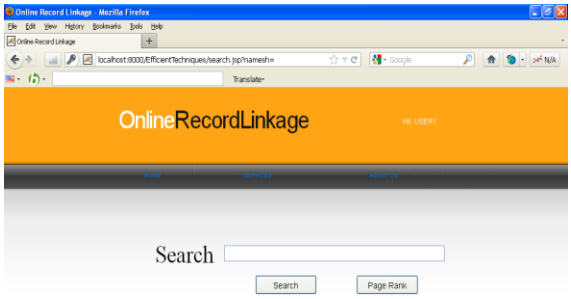
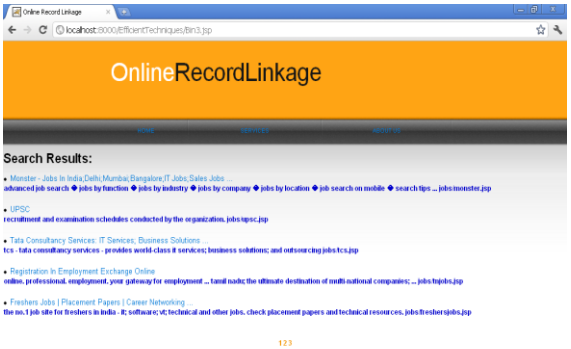
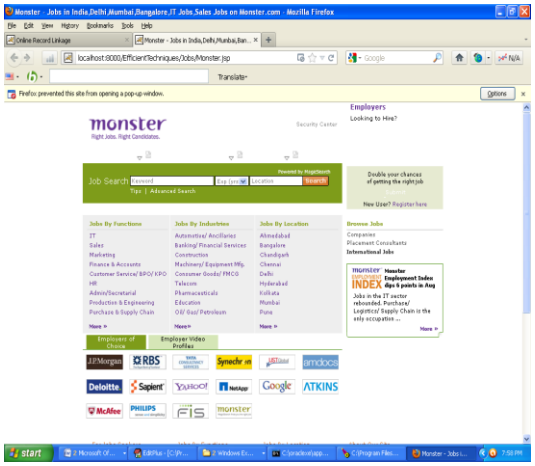
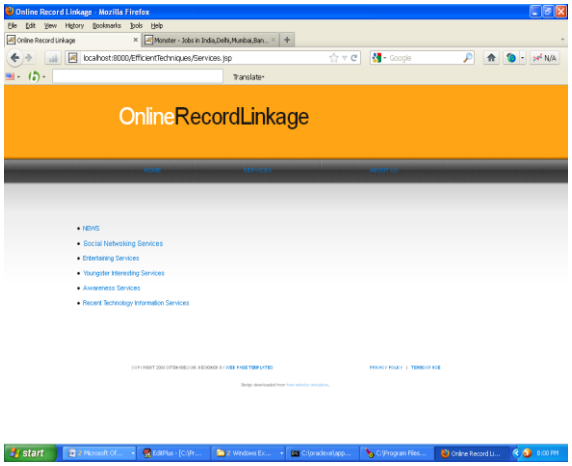


Fig: User Registration Page

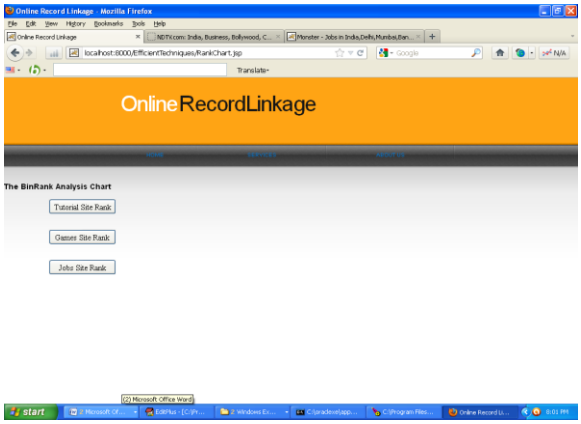

OnlineRecordLinkage
Search
Search
Page Rank

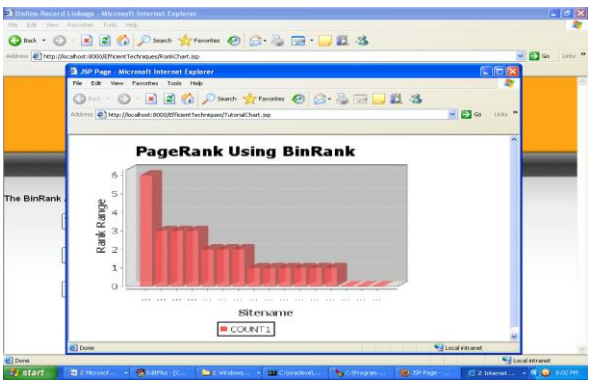

OnlineRecordLinkage
Search Results:


monster


OnlineRecordLinkage


NDTV
HTTP Status 404 -
/EfficientTechniques
/Services
/News/NDTV


OnlineRecordLinkage
The BinRank Analysis Chart
Tutorial Site Rank
Games Site Rank
John Site Rank


PageRank Using BinRank
Rank Range
Sitename
COUNT(1)

IV. Conclusion:

Efficient techniques to facilitate record linkage decisions in a distributed, online setting. Record linkage is an important issue in heterogeneous database systems where the records representing the same real-world entity type are identified using different identifiers in different databases. In the absence of a common identifier, it is often difficult to find records in a remote database that are similar to a local enquiry record. Traditional record linkage uses a probability-based model to identify the closeness between records. The matching probability is computed based on common attribute values. This, of course, requires that common attribute values of all the remote records be transferred to the local site. The communication overhead is significantly large for such an operation. Propose techniques for record linkage that draw upon previous work in sequential decision making.More specifically, we develop a matching tree for attribute acquisition and propose three different schemes of using this tree for record linkage.

Future Enhancement

Although this work presents a better approach for version differentiation to select the exact updated coverage from old version to new version by using the statistical information, we can still improve by automating without selecting the updated coverage manually by the test engineer. By this we can still reduce the testing budget and maintenance.

In regression testing many techniques are proposed that provides an efficient way of selecting regression test suite without rerunning all test cases that was developed for old version. But still IT companies are not utilizing these techniques, because these techniques are not assuring completely and test engineers are still using manual testing and automation testing for safety. I can assure that my proposed approach with enhancement is safe and precise in that it computes exactly the same information as if all test cases in the test suite were rerun.

References

[1] J.A. Baldwin, "Linked Record Health Data Systems," The Statistician, vol. 21, no. 4, pp. 325-338, 1972.
[2] C. Batini, M. Lenzerini, and S.B. Navathe, "A Comparative Analysis of Methodologies for Database Schema Integration,"ACM Computing Surveys, vol. 18, no. 4, pp. 323-364, 1986.
[3] R. Baxter, P. Christen, and T. Churches, "A Comparison of Fast Blocking Methods for Record Linkage," Proc. ACM Workshop Data Cleaning, Record Linkage and Object Consolidation, pp. 25-27, Aug.2003.
[4] T.R. Belin and D.B. Rubin, "A Method for Calibrating False-Match Rates in Record Linkage," J. Am. Statistical Assoc., vol. 90, no. 430,pp. 694-707, 1995.
[5] P. Bernstein, "Applying Model Management to Classical Meta Data Problems," Proc. Conf. Innovative Database Research (CIDR),pp. 209-220, Jan. 2003.

Books

1. Java 2 complete Reference by Herbert Schieldt
2. "Software Engineering", A Practitioner's Approach, 6th Edition, Tata McGrawHill
3. "Software Testing principles and practices", Srinivasan Desikan, Gopala swamiRamesh, Pearson edition, India.
4. A Unified Modeling Language User Guide, 2nd edition, Book by Grady Booch, James RamBaugh, IvarJacobson for UML concepts and models.

Websites

1.www.google.co.in/WirelessMeshNetwork.doc
2.www.wikiedia/Network-Security.com
3.www.ieeeexplore.ieee.org/xpl/anonymous/networks.pdf

IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 11, Issue 4 (May. - Jun. 2013), PP 44-47
www.iosrjournals.org

Crypt Sequence DNA

Mrs.S.Sujatha MCA, MPhil.,(Ph.D.),Bhadra Prabhakaran
School of IT and Science,Dr.G.R. Damodaran College of Science, India

Abstract: *The DNA sequence plays a major role in identifying each individual and making them unique. So we try incorporating DNA to encrypt the data in exclusively identifying individual encryption format. This method would highlight very basic and easy algorithms to form exclusive cipher text for each data or file. Each of the DNA component would be allocated a fixed algorithm such that the encryption would be based on the components sequence without altering the algorithms fixed.*
Keywords *-Cipher text, Component sequence DNA, DNA Component and Encrypt*

I. INTRODUCTION

Securing data is the massive issue in today's world. Human beings are the source of data. The data obtained from or created by them is always insecure in the way of storage. The data storage was made secured by cryptography principles.Data in any format or the source has the knack to be publicized due to the advancement in the cryptanalysis. This work involves another security measure of data. Any form of cryptography for a certain amount of data would have same style of encryption throughout the entire domain. This work proposes an individual format for every data present in the data domain. The individuality needed a specific itinerary and that was found as the DNA of the individual. DNA is a unique composition of molecules for every human who is the source of data. The sequence of DNA is the key to the complete encryption method.

II. LITERATURE REVIEW

DNA is used only for the encrypted storage of data in a tiny memory according to the researchers A new method of data storage that converts information into DNA sequences allows you to store the contents of an entire computer hard-drive on a gram's worth of E. coli bacteria[1]. The DNA composition sequence is faked with the encrypted sequence of data as the researchers showed how to change the word "iGEM" into DNA-ready code. They used the ASCII table to convert each of the individual letters into a numerical value [2]. Binary data is encoded in the geometry of DNA nanostructures with two distinct conformations. Removing or leaving out a single component reduces these structures to an encrypted solution [3].Alteration of DNA sequence using one time pad and adding the encrypted message into the original DNA sequence strand.[4]The keys need to be binary string and the image file input has to be converted to binary data[5].DNA based Cryptography which puts an argument forward that the high level computational ability and incredibly compact information storage media of DNA computing has the possibility of DNA based cryptography based on one time pads[6]. An encryption scheme is designed by using the technologies of DNA synthesis, PCR amplification and DNA digital coding as well as the theory of traditional cryptography [7].

III. PROBLEM DEFINITION

The database is still not secure with the encrypted form of data.The keys or even the plain texts are identified by cryptanalysis like Brute Force. There is no specific individual encryption format for every specific data that is stored in the database.The individual encryption is not possible in huge database. Every data cannot be encrypted individually. This is the reason why hacking becomes stress-free all times and if one part of the encryption algorithm is acknowledged then the whole database can be scythed.

IV. PROPOSED WORK

First step in this search, Chargaff set out to see whether there were any differences in DNA among different species. After developing a new paper chromatography method for separating and identifying small amounts of organic material, Chargaff reached two major conclusions (Chargaff, 1950). First, he noted that the nucleotide composition of DNA varies among species. In other words, the same nucleotides do not repeat in the same order, as proposed by Levene. Second, Chargaff concluded that almost all DNA--no matter what organism or tissue type it comes from--maintains certain properties, even as its composition varies. In particular, the amount of adenine (A) is usually similar to the amount of thymine (T), and the amount of guanine (G) usually approximates the amount of cytosine (C). In other words, the total amount of purines (A + G) and the total amount of pyrimidines (C + T) are usually nearly equal. (This second major conclusion is now known as "Chargaff's rule.") Chargaff's research was vital to the later work of Watson and Crick, but Chargaff himself could not imagine the explanation of these relationships--specifically, that A bound to T and C bound to G

within the molecular structure of DNA. Chargaff's realization that A = T and C = G, combined with some crucially important X-ray crystallography work by English researchers Rosalind Franklin and Maurice Wilkins, contributed to Watson and Crick's derivation of the three-dimensional, double-helical model for the structure of DNA. Watson and Crick's discovery was also made possible by recent advances in model building, or the assembly of possible three-dimensional structures based upon known molecular distances and bond angles, a technique advanced by American biochemist Linus Pauling. In fact, Watson and Crick were worried that they would be "scooped" by Pauling, who proposed a different model for the three-dimensional structure of DNA just months before they did. In the end, however, Pauling's prediction was incorrect.

Using cardboard cutouts representing the individual chemical components of the four bases and other nucleotide subunits, Watson and Crick shifted molecules around on their desktops, as though putting together a puzzle. They were misled for a while by an erroneous understanding of how the different elements in thymine and guanine (specifically, the carbon, nitrogen, hydrogen, and oxygen rings) were configured. Only upon the suggestion of American scientist Jerry Donohue did Watson decide to make new cardboard cutouts of the two bases, to see if perhaps a different atomic configuration would make a difference. It did. Not only did the complementary bases now fit together perfectly (i.e., A with T and C with G), with each pair held together by hydrogen bonds, but the structure also reflected Chargaff's rule.

Deoxyribonucleic acid (**DNA**) is a molecule that encodes the genetic instructions used in the development and functioning of all known living organisms and many viruses. Along with RNA and proteins, DNA is one of the three major macromolecules essential for all known forms of life. Genetic information is encoded as a sequence of nucleotides (guanine, adenine, thymine, and cytosine) recorded using the letters G, A, T, and C. Most DNA molecules are double-stranded helices, consisting of two long polymers of simple units called nucleotides, molecules with backbones made of alternating sugars (deoxyribose) and phosphate groups (related to phosphoric acid), with the nucleobases (G, A, T, C) attached to the sugars. DNA is well-suited for biological information storage, since the DNA backbone is resistant to cleavage and the double-stranded structure provides the molecule with a built-in duplicate of the encoded information.

These two strands run in opposite directions to each other and are therefore anti-parallel, one backbone being 3' (three prime) and the other 5' (five prime). This refers to the direction the 3rd and 5th carbon on the sugar molecule is facing. Attached to each sugar is one of four types of molecules called nucleobases (informally, *bases*). It is the sequence of these four nucleobases along the backbone that encodes information. This information is read using the genetic code, which specifies the sequence of the amino acids within proteins. The code is read by copying stretches of DNA into the related nucleic acid RNA in a process called transcription.

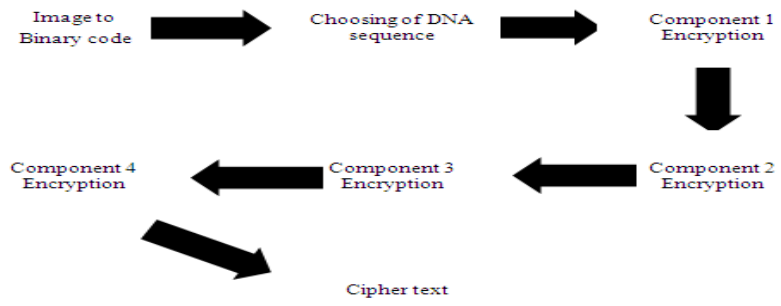
These four compositions can be provided with specific fixed algorithms and the encryption sequence would be based on the DNA composition sequence.The algorithm would be fixed for each individual component. The encryption sequence would be based on the sequence of the components present in the individual's DNA. This would lead to reduction of complicate algorithms and also a lot of keys to be remembered. This makes data so secure that even if the keys are identified the plain text could not be retrieved.

The DNA composition consists of four major components

- Adenine
- Thymine
- Guanine
- Cytosine

These four compositions can be provided with specific fixed algorithms and the encryption sequence would be based on the DNA composition sequence.The algorithm would be fixed for each individual component. The encryption sequence would be based on the sequence of the components present in the individual's DNA. This would lead to reduction of complicate algorithms and also a lot of keys to be remembered. This makes data so secure that even if the keys are identified the plain text could not be retrieved.

V. METHODOLOGY



VI. Algorithms Chosen And Their Compatibility

	AES	Triple DES	Blow Fish	Serpent
Derived from	Square	DES	-	Square
Key sizes	128, 192 or 256 bits	168, 112 or 56 bits	32–448 bits	128, 192 or 256 bits
Block sizes	128 bits	64 bits	64 bits	128 bits
Structure	Substitution-permutation network	Feistel network	Feistel network	Substitution-permutation network
Rounds	10, 12 or 14 (depending on key size)	48 DES-equivalent rounds	16	32
cryptanalysis	Brute Force, Linear, Boomerang and Differential	Brute Force, Linear, Boomerang and Differential	Brute Force, Linear, Boomerang and Differential	Brute Force, Linear, Boomerang and Differential
Security	All known attacks are computationally infeasible. For AES-128, the key can be recovered with a computational complexity of $2^{126.1}$ using bicliques. For biclique attacks on AES-192 and AES-256, the computational complexities of $2^{189.7}$ and $2^{254.4}$ respectively apply.	The original DES cipher's key size of 56 bits was sufficient but the computational power made brute-force attacks feasible. Triple DES provides a method of increasing the key size of DES making it stronger without changing cipher blocks	Four rounds of Blowfish are susceptible to a second-order differential attack but when the keys are still stronger then it becomes strong enough	All known attacks are computationally infeasible. An attack in 2011 could break only till 12 Round of the Serpent decryption

VII WORKING PRINCIPLE

1. The image file is converted to binary string and keys are generated
2. DNA sequence is chosen from the complete sequence
3. Encryption of data according to the DNA component sequence chosen
4. Adenine- AES Algorithm
5. Thymine- Triple DES
6. Cytosine- Serpent Algorithm
7. Guanine- Blowfish Algorithm

VIII PERFORMANCE

- Produces a very complex cipher code such that it is hard to break
- The algorithms are unique in their keys and also their encryption format which makes cryptanalysis more stressful
- Cipher code can be identified or cracked only when the DNA random generated sequence components are found with the keys and the algorithms allocated to them, which would be really hectic.

IX CONCLUSION

This method would help in encrypting the data individually though all the data belong to the same database. The common encryption format for a database can be totally eradicated. The security would be at a very high stage. Government confidential material can be at a higher rate of confidentiality and in very simple steps.

REFERENCES

Journal Papers:

1)Bio encryption can store almost a million gigabytes of data inside bacteria-*IO9 Journal Vol 2 by Alasdair Wilkins* on Nov 26 2010 2)Binary DNA Nanostructures for Data Encryption- *PLOS ONE* Published: September 11, 2012 by *Ken Halvorsen, Wesley P. Wong* 3)DNA-Based Data Encryption and Hiding Using Play fair and Insertion Techniques- *Journal of Communications and Computer Engineering Vol2, 3 (2012) by A Atito* 4)DNA-based Cryptography -DNA Based Computers- *June 1999 by Ashish Gehani, Thomas H. LaBean and John H. Reif* 7) An encryption scheme using DNA technology Guangzhao Cui Coll. of Electr. Inf. Eng., *Zhengzhou Univ. of Light Ind., Oct. 1 2008,*

Books:

Advanced Encryption standard Algorithm by VincentRijmen, Joan Daemen, Triple Data Encryption Standard, Serpent Algorithm by Ross Anderson, Eli Biham, Lars Knudsen, Blow Fish Algorithm, Text book on Identity-Based Encryption By Chatterjee Sanjit, SarkarPalash, A text book A Novel DNA based Encrypted Text Compression By D. Prabhu, M. Adimoolam, P.Saravannan, A text book DNA Structure and Function By Richard R Siden,

Articles:

ARTICLE ON INNOVATIVE FIELD OF CRYPTOGRAPHY: DNA CRYPTOGRAPHY BY ERSONI – 2012 ER.RANUSONI, ER.VISHAKHASONI AND ER.SANDEEPKUMARMATHARIYA, ARTICLE ON STABILIZING SYNTHETIC DATA IN THE DNA OF LIVING ORGANISMS SYST SYNTH BIOL. 2008 JUNE ,NOZOMUYACHIE, YOSHIAKIOHASHI, AND MASARUTOMITA

IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 11, Issue 4 (May. - Jun. 2013), PP 48-53
www.iosrjournals.org

A Novel Algorithm for Mining Fuzzy High Utility Itemset from Fuzzy Transaction Database

Mrs. Monisha M.[1], Mrs. Mala A.[2] , Dr. F. Ramesh Dhanaseelan[3]
[1]*M.E-(Final Year) Department of Computer Science and Engineering, PSN College of Engineering and Technology, Tirunelveli, India*
[2] *Associate Professor, Department of Computer Science and Engineering, PSN College of Engineering and Technology, Tirunelveli, India.*
[3]*Prof. & Head, Department of Computer Applications, St. Xavier's Catholic College of Engineering, Nagercoil*

Abstract: *Utility mining discovers the most profitable item in transaction database which is of great importance in the upgrade of revenue. There are various algorithms for mining the high utility itemset such as CTU-PRO, Two phase and also UMMI algorithm. These algorithms mine the high utility itemset in an efficient way, but they do not reflect the fuzzy degree of purchased quantity which is essential for making decision in various applications like sales analysis and stock control. Itemsets with profit slightly less than the threshold is also discarded. To overcome these problems, fuzzy set theory is applied to the utility mining problem and a novel algorithm namely fuzzy high utility item-mine (FHUI-Mine) is introduced to mine the fuzzy high utility itemset. The quantity information from transaction database is fuzzified in order to reflect the fuzzy degree of purchased quantity. FHUI-Mine also provides a fuzzy threshold range that may include the itemsets whose support is slightly less than the designated threshold. More over this algorithm also provides the utility information due to fuzzification of threshold. To prove the feasibility of FHUI-Mine, it was compared with the well known UMMI algorithm through experimental evaluation. The results show that FHUI-Mine delivers higher mining capability as it can not only mine all high utility itemset but also discover additional itemsets that are potentially high utility ones.*

Keywords: *Data Mining, Fuzzy Mining, Fuzzy set, Fuzzification, Utility Mining,*

I. Introduction

Data mining is the process of extracting hidden patterns from the database that is potentially useful for decision makers to gain knowledge and sales analysis. One of the important tasks in data mining is utility mining which refers to the discovery of more profitable item. High utility mining mines the high utility itemset from the transaction database. When the utility of an item is greater than or equal to user specified minimum utility threshold, then it is a high utility item. For example, assume the frequency of item A is 7, item B is 6, and itemset AB is 3. The profit of item A is 2, B is 5. The utility value of A is 7 * 2 = 14, B is 6* 5= 30, and AB is 3 * 2 + 3 * 5 = 21. If the minimum utility threshold is 25, B is a high utility itemset

A number of algorithms have been proposed for high utility itemset mining namely Two phase [1], TWU mining [2], UMMI algorithm [3]. Two phase algorithm mines the high utility itemset in two phases using the transaction weighted downward closure property in phase I to find the HTWU (High Transaction Weighted Utility) item. Phase II mines the high utility itemset from phase I output of HTWU item. TWU mining use a tree structure to capture the utility information. UMMI algorithm is introduced to overcome the shortcoming of two phase, TWU mining which consists of large number of HTWU itemsets thereby increasing the execution time. High utility itemsets can be mined using two phases: maximal phase and utility phase. Maximal phase mines maximal transaction weighted utility (MTWU) item using maximal itemset property and utility phase discovers the high utility itemset from MTWU item using Mlex tree.

However, there exists some weakness in the existing algorithm for high utility itemset mining: 1.Itemsets with profit slightly less than the designated threshold value is discarded. For example, if the minimum utility threshold is 130, itemsets with profit 129 will be pruned even though the itemset may be significant. 2. Quantity information about the high utility itemset is not reflected that a sales manager is interested in. 3. Profit degree information is also not provided in the existing algorithm. These shortcomings affect the ability to select the more profitable and cost-saving products.

The weakness of high utility itemset mining can be overcome by applying fuzzy set theory. Fuzzy set theory yields better results when applied in data mining. Fuzzy logic is mainly used to determine the uncertainty of particular item in which the membership value lies in the interval [0, 1].

A novel method namely FHUI (Fuzzy High Utility Itemsets)-*Mine* is proposed which can reflect the quantity information and also profit information. In this algorithm, itemsets with profit slightly less than the designated threshold value is also included resulting in new revenue opportunities. FHUI-Mine consists of two

phases. In phase I, a fuzzy membership function is defined to represent the quantities in fuzzy sets. Thus the transaction table is transformed into fuzzy transaction table. Then fuzzy transaction utility will be calculated from the fuzzy transaction table and utility table. Further, Fuzzy transaction weighted utility is produced to discover the phase I high utility itemset. Phase II calculates the fuzzy utility for the phase I high utility itemset to yield the fuzzy high utility itemset.

In order to prove the feasibility of proposed algorithm, FHUI-Mine is compared with UMMI algorithm for the same threshold. Through experimental evaluation, it is proven that FHUI-Mine algorithm can not only discover the high utility itemset but also detect itemsets with profit slightly less than the designated threshold. Moreover, it provides the important quantity and profit information of high utility itemset through fuzzification.

1.1 Paper Organization

The rest of the paper is organized as follows. Section 2 narrates about membership function, fuzzy quantity, fuzzy utility, fuzzy transaction utility and fuzzy transaction weighted utility. Section 3 describes about the proposed algorithm FHUI-Mine. Section 4 explains about experimental results and the paper is concluded in Section 4

II. Preliminaries

2.1 Membership function

The first essential thing to transform transaction database to fuzzy transaction database is to define a membership function for quantity shown in figure 1 which can transform a quantity value in to fuzzy quantity membership value and fuzzy membership region. This fuzzification of quantity reflects the fuzzy degree of purchased quantity thereby providing additional quantity information. Here the membership function used is triangular. There are three regions of quantity namely low, middle and high. Region low is defined between the interval 0-6; region middle is defined between the interval 1-11 and region high was between 11-∞ interval.

Membership value is always defined between [0, 1] and the fuzzy transaction database for quantity [4] lies in the value 0 to 1. X axis represent the membership value and Y axis represent the number of item. The membership function for triangle can be define by a lower limit a, an upper limit b and a value m, where a<m<b.

$$\mu_{A(x)} = \begin{cases} 0, & x \le a \\ x-a/m-a, & a < x \le m \\ b-x/b-m, & m < x < b \\ 0, & x \ge b \end{cases} \quad \textbf{(1)}$$

Where $\mu_{A(x)}$ represents the grade of membership of the element x to fuzzy set A.

2.2 Fuzzy quantity

The equation to find the fuzzy quantity value of item i_p in transaction T_q is denoted as $fo\ (i_p, T_q)$ which is defined as,

$$fo\ (i_p, T_q) = \sum fq\ (i_p, T_q, j) \times weight\ (j) \quad \textbf{(2)}$$

Where $fq\ (i_p, T_q, j)$ is the fuzzy value of fuzzy region j and *weight (j)* is a variable parameter defined by the fuzzy region. If a fuzzy region is low then the weight should be low when compared to the region middle and high. Special weights are assigned for region low, medium and high.

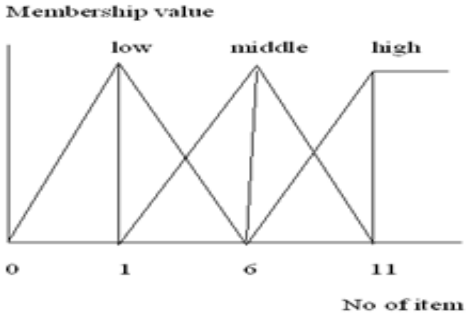


Figure 1: The fuzzy membership function for quantity

2.3 Fuzzy utility

In high utility itemset mining, item utility is equal to quantity value multiplied by profit. In fuzzy transaction, the similar procedure to find item utility will produce false results. For example, item A occurs in different transaction T1 and T5. If the fuzzy set of (A, T1) = {1/L, 0/M, 1/H} and (A, T5) = {0/L, 0/M, 1/H}. If

their item utilities are obtained by multiplying fuzzy quantity with profit, item utilities of T1 and T5 will be same even though T5 yields a higher quantity value. The fuzzy utility is defined as,

$$fu(i_{p,}T_{q)} = fo\ (i_{p,}T_{q)} \times s(i_p) \qquad \textbf{(3)}$$

Where $s(i_p)$ is the value associated with item i_p in utility table

2.4 Fuzzy transaction utility

Fuzzy transaction utility can be defined as the sum of the fuzzy utilities of item occurring in the particular transaction. The equation denoting fuzzy transaction utility is

$$ftu\ (T_q) = \sum_{i_p \varepsilon T_q} fu\ (i_{p,}T_{q)} \qquad \textbf{(4)}$$

2.5 Fuzzy transaction weighted utility

It is the sum of fuzzy transaction utilities of item occurring in the particular transaction for an item. The equation denoting fuzzy transaction weighted utility can be defined as,

$$ftwu\ (x) = \sum_{xcT_q \varepsilon D} ftu\ (T_{q)} \qquad \textbf{(5)}$$

III. Proposed Algorithm: Fhui-Mine

Fuzzy high utility itemset mining algorithm is proposed to provide more information concerning the quantity and profit information of high utility itemset. FHUI-Mine algorithm is intended for mining fuzzy high utility itemset by applying fuzzy theory to high utility itemset. The algorithm comprises of two phases: phase I mines High Fuzzy Transaction Weighted Utilization Itemsets and phase II mine fuzzy high utility itemsets from phase I high utility itemset.

3.1 Phase I: Mining High Fuzzy Transaction-Weighted Utilization Itemset

In phase I, first a fuzzy membership function for quantity is defined to provide the fuzzy quantity region which is shown in figure 1. It can transform a quantity value in to fuzzy membership region and membership value thereby enhancing transaction database as fuzzy transaction database. The transaction database is shown in table 1.

Table 1: Transaction database

	A	**B**	**C**	**D**	**E**
T01	0	3	6	1	4
T02	3	0	10	0	9
T03	7	0	4	0	0
T04	6	1	0	1	0
T05	0	0	8	0	3
T06	0	2	12	1	0
T07	9	0	0	0	7
T08	2	2	0	0	6

The quantity attribute for fuzzy membership function has three fuzzy regions namely low, middle and high. Thus, fuzzy membership value for the purchased quantity is represented as fuzzy set in terms of {fuzzy value of low/low, fuzzy value of middle/middle, and fuzzy value of high/high}. For example, the quantity value '9' is converted in to fuzzy set as {0.0/L, 0.4/M, 0.6/H}, where 'L', 'M' and 'H' are acronym of 'Low', 'Middle' and 'High'.

To find the sole representing fuzzy quantity from three fuzzy regions, a maximum value is generated from the fuzzy set. For example, the sole representation for quantity value '9' is represented as *max* (0.0, 0.4, 0.6) =0.6. The sole representation for the quantity value was done according to equation (2). Weight parameter was fixed for the three region as Low=0.1, Middle=0.5, High=1. According to equation (2), sole representation for quantity '9' is 0.6. The calculation is preceded as follows. The max value in fuzzy set is 0.6 which is present in region high. The max value is multiplied with weight assigned for region high to yield the sole value for particular quantity. Hence this procedure is adopted for all quantity in transaction database to reproduce fuzzy transaction database which is shown in table 2.

Table 2: Fuzzy Transaction database

	A	**B**	**C**	**D**	**E**
T01	0	0.06	0.5	0.1	0.3
T02	0.06	0	0.8	0	0.6
T03	0.4	0	0.3	0	0
T04	0.5	0.1	0	0.1	0
T05	0	0	0.3	0	0.06
T06	0	0.08	0.9	0.1	0
T07	0.6	0	0	0	0.4
T08	0.08	0.08	0	0	0.5

Table 3: Utility table

Item	A	B	C	D	E
Utility	5	11	3	20	4

After the transformation of fuzzy transaction database from transaction database, the next step is to find the fuzzy utility for each and every transaction according to equation (3) by multiplying fuzzy quantity with profit value in utility table shown in table 3. Fuzzy utility for the item 'B' in transaction T01 is defined as the quantity value of item 'B' is 0.06 which is multiplied by the profit value for 'B' in the utility table which is 11 to yield the fuzzy utility as 0.66. Fuzzy utility values for the corresponding transaction databases are displayed in table 4.Fuzzy utility calculated was used to find fuzzy transaction utility Fuzzy according to equation (4) for each and every transaction. Fuzzy transaction utility table is displayed in table 5

Table 4: Utility table

	A	**B**	**C**	**D**	**E**
T01	0	0.66	1.5	2	1.2
T02	0.3	0	2.4	0	2.4
T03	2	0	0.9	0	0
T04	2.5	1.1	0	2	0
T05	0	0	0.9	0	0.24
T06	0	0.88	2.7	2	0
T07	3	0	0	0	1.6
T08	0.4	0.88	0	0	2

Table 5: Fuzzy transaction utility

Transaction ID	**Transaction Utility**
T01	5.36
T02	5.1
T03	2.9
T04	5.6
T05	1.14
T06	5.58
T07	4.6
T08	3.28

The total database utility is 30.86 which is the sum of transaction utility of all the transaction. The threshold was set to 25% which then yields the minimum utility threshold as,

Min_utility threshold=Total database Utility× threshold

=33.56×25%

=8.39

Fuzzy transaction weighted utility itemset was then found from fuzzy utility according to equation (5). Fuzzy TWU for item 'A' has been evaluated as,

Ftwu (A) = T02+T03+T04+T07+T08

=5.1+2.9+5.6+4.6+3.28

=21.48

Ftwu (A) is high fuzzy transaction weighted utility item since the ftwu(A) is greater than the min.utility threshold. Similarly for each item and itemset fuzzy transaction weighted utility has been evaluated to find whether it is high fuzzy transaction weighted utility item. Hence phase I high utility itemsets are obtained as,

{A}, {B}, {C}, {D}, {E}, {AB}, {AE}, {BC}, {BD}, {BE}, {CD}, {CE}, {BCD}.

3.1 Phase II: Mining fuzzy high utility itemset

In phase II, high utility itemets obtained in phase I are scanned to find the fuzzy high utility itemset which is greater than the min. utility threshold. The utility of item 'C' is calculated as, {C}=1.5+2.4+0.9+0.9+2.7=8.4. Fuzzy high utility items produced in phase II are, {C-8.4}, {AE-9.7}, {BD-8.64}, {CE-8.64}, {BCD-9.74}

The utility information for the mined fuzzy high utility itemsets is defined by representing the utility value in to utility region. The utility of each phase II high utility itemsets are fuzzified in to three regions namely middle, high and very high as shown in figure 2.

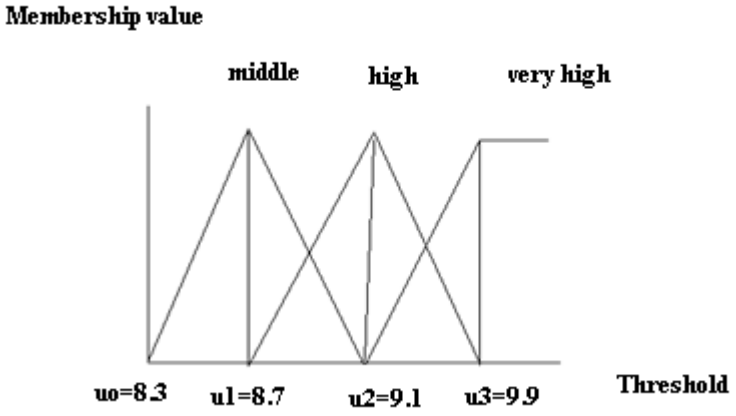


Figure 2: Fuzzified utility

The values for u_0**,** u_1, u_2, u_3, u_4 are predefined as u_0=8.3**,** u_1=1.05* u_0, u_2=1.1* u_0, u_3=1.2* u_0, where u_0 is the min.utility threshold. For the fuzzy high utility itemsets obtained, the utility information can be hence obtained.

IV. Experimental Results

Fuzzy high utility itemsets are mined using FHUI-mine algorithm and high utility itemsets are mined using UMMI algorithm. The performance of FHUI-mine algorithm has been evaluated using synthetic datasets. The quantity of each item in each transaction is taken randomly and the utility tables are also created synthetically by taking the utility values in random manner. Number of transaction is set to 8(T1-T8) containing 5 items (A, B, C, D, and E). An item may or may not occur in the transaction and each transaction atleast contains some of the items. For both FHUI-Mine and UMMI, threshold value was set to 25%. In high utility mining ordinary transaction database was used and in fuzzy high utility mining , fuzzy transaction database has been created from transaction database. Quantity values are fuzzified in to low, middle and high region and threshold values are fuzzified in to middle, high and very high region.

In UMMI algorithm, the phase-I itemsets are obtained as{AE}, {BE}, {C E}, {BCD} which is obtained using maximal itemset property where as in FHUI-Mine phase-I high utility itemsets are {A}, {B}, {C}, {D}, {E}, {AB}, {AE}, {BC}, {BD}, {BE}, {CD}, {CE}, {BCD} from transaction weighted utility property. In phase-II UMMI produced {A}, {AE}, {CE}, {BCD} from Mlex tree whereas in FHUI- mine, the high utility itemsets produced are {C-8.4}, {AE-9.7},{BD-8.64},{CE-8.64}, {BCD-9.7} by calculating the utility. FHUI-Mine discovers all the itemsets mined by UMMI algorithm and also discovers additional itemsets.

As fuzzification is not used in high utility mining, the high utility information about the item is not provided and items with nearest threshold value will also be discarded. In FHUI mine the fuzzy high utility itemsets obtained are present in very high region and hence it will be more important for the decision makers in company to attain more profit and less experts will also gain more knowledge about yielding more profit. Hence Fuzzy high utility mining provides more information about the utility of an item and items with nearest threshold value was also not discarded due to fuzzification.

Table 5: UMMI algorithm mining results

High utility Itemset	**Utility**
{A}	135
{AE}	158
{CE}	164
{BCD}	149

Table 6: FHUI-Mine algorithm mining results

Fuzzy High utility Itemset	Fuzzy Utility	Fuzzy utility set
{C}	8.4	{0.2/M,0/H,0/VH}
{AE}	9.7	{0/M,0.2/H,0.7/VH}
{BD}	8.64	{0.7/M,0/H,0/VH}
{CE}	8.64	{0.7/M,0/H,0/VH}
{BCD}	9.7	{0/M,0.2/H,0.7/VH}

In table 6, 'M', 'H', 'VH' are the acronym of Middle, High and Very High. In UMMI algorithm, utility of items {C}, {BD } is discarded eventhough the utility is nearest to the minimum utility threshold which came in to existence in FHUI-Mine algorithm. Through experimental evaluation, mining results shows that the proposed FHUI-Mine reflects the fuzzy degree of purchased quantity and profit of high utility itemsets and the purchased frequency of each resulting high utility with combination of fuzzy quantity regions that stores are interested.

V. Conclusion

The proposed FHUI-Mine provides a narrow expansion of high utility itemset mining. Phase I mines the high utility itemset and phase II mines the fuzzy high utility itemset. Itemsets with profit slightly less than the designated threshold is not discarded. Moreover, it also provides the information regarding fuzzy degree of purchased quantity and also utility information. In real applications, it allows decision makers to gain unprecedented insight into all aspects of their business by offering them to fuzzified and maximize the profit they can make. The utility degree information reflects the shop's profitability and also it provides the ability to select the most lucrative and cost saving products and also used for sales analysis.

References

[1] AlokChoudhary, Ying Liu, Wei-keng Liao, (2005), 'A Fast High Utility Itemsets Mining Algorithm', *UBDM'05 Chicago, USA*

[2] Bay VO, Huy N guyen, Bac Le (2009) 'Mining High Utility Itemset from Vertical Distributed Database', *IEEE Xplore.*

[3] Ming-Yen Lin, Tzer-Fu Tu, Sue-chen Hsueh, 'High utility pattern mining using the maximal itemset property and lexicographic tree structures, *Science Direct, journal of information sciences* 215(2012)1-14.

[4] Tzung-Pei Hong, Kuei-Ying Lin, Shyue-Liang Wang, 'Fuzzy data mining for interesting generalized association rules', *Elsevier on fuzzy sets and system* 138(2003) 255-269.

[5] R. Agrawal, T. Imielinskl, A. Swami, 'Mining Association rules between sets of items in large databases', *Proceedings of the 1993 ACM SIGMOD International Conference in Management of Data, Washington, DC*, 1993, pp. 207–216.

[6] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: Proceedings of 20th International Conference on Very Large Databases, Santiago, Chile, 1994, pp. 487–499.

[7] Alva Erwin,Raj p. Gopalan, N.R.Achunthan (2007), 'A Bottom-Up Projection Based Algorithm For mining High Utility Itemset', *Workshop on Integrating AI and Data Mining (AIDM), Australia*,Vol.84.

[8] Chia-Ming Wang, Shyh-Huei Chen; Yin-Fu Huang, 'A fuzzy approach for mining high utility quantitative itemsets', *Fuzzy Systems, 2009. FUZZ-IEEE* 2009.

[9] T.P. Hong, C.Y. Lee, Induction of fuzzy rules and membership functions from training examples, *Fuzzy Sets and Systems* 84 (1996) 33–47.

[10] Sandeep Kumar Singh, Mr.Ganesh Wayal, Mr.Niresh sharma, 'A Review: Data Mining with Fuzzy Association Rule Mining', *International Journal of Engineering Research & Technology (IJERT)* Vol. 1 Issue 5, July – 2012.

[11] Stergios Papadimitriou Seferina Mavroudi, 'The Fuzzy Frequent Pattern Tree'.

[12] Karthikeyan T, Samuel Chellathurai A and Praburaj B, 'A study on a novel method of mining fuzzy association using fuzzy correlation analysis', *African Journal of Mathematics and Computer Science Research* Vol. 5(2), pp. 28-33, 15 January, 2012.

[13] H. Yao, H.J. Hamilton, 'Mining itemset utilities from transaction databases', *Data & Knowledge Engineering* 59 (3) (2006) 603–626.

[14] Vincent S. Tseng and C. P. Kao, 'A Novel Similarity-based Fuzzy Clustering Algorithm by Integrating PCM and Mountain Method', *IEEE Transactions on Fuzzy Systems*, vol. 15, Issue 6, pp. 1188-1196.

IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 11, Issue 4 (May. - Jun. 2013), PP 65-69
www.iosrjournals.org

Evaluating Air Pollution Parameters Using Zigbee (IEEE 802.15.4)

Darshana N. Tambe[1], Nekita A. Chavhan[2]
[1] *Wireless Communication & Computing, Department of Computer Science & Engineering, G.H. Raisoni College of Engineering, Nagpur, India.*
[2] *Department of Computer Science & Engineering, G. H. Raisoni College of Engineering, Nagpur, India*

Abstract: *Air pollution receives one of the prime concerns in India, primarily due to rapid economic growth, industrialization and urbanization with associated increase in energy demands. Lacks of implementation of environmental regulations are contributing to the bad air quality of most of the Indian cities. Air pollutants produced in any air shed are not completely confined, but at time passing all the geographical boundaries, hence donot remain only a problem of urban centers, but spread and affect remote rural areas supporting large productive agricultural land. In environmental parameters the air pollution is measure by taking one or few samples in a day that means there is no information present about the real time air pollution data. This is the main disadvantages of such system. Most of the countries in the world work on the real time bases to monitor the air quality. In this paper we describe use of ZigBee, sensor nodes, GPS to construct distributed system for urban air pollution monitoring and control. ZigBee module and pollution server is interfaced with GPS system to display real-time pollutants levels and there location on a 24h/7 days basis. In this system there are four transmitter (Node1, Node2, Node3, Node4) are present which transmit the different levels of pollutant substance such as CO2, SO2, and NO2 to the receiver node in real time. The system was successfully tested in the G.H. Raisoni College of Engineering, Nagpur, India.*

Keywords: *ZigBee Sensor Node, Air Pollution, GPS, Environmental Pollution, Real Tim, CO2, SO2, NO2.*

I. Introduction

With fast development of the industrialization and urbanization process in the world, environmental pollution problems become more universal. At present environment contains air pollution, water pollution and soil pollution worldwide. Air pollution is the presence of contaminants or pollutant substances in the air that interfere with human health or welfare, or produce other harmful environmental effects. The World Health Organization states that 2.4 million people pass away each year because of air pollution. Based on the fact above mentioned, the human should focus on design air pollution monitoring method.

Pure air and human health goes Hand in Hand. Air pollution is harmful for human Health. It causes difficulty in breathing, wheezing, coughing and many respiratory problems. Currently there are two methods to monitor air pollution at present. First one is non automatic and other is automatic. The advantages of non automatic sampling method are monitoring devices is simple and inexpensive but it monitors the parameters for certain period. It does not provide the real time monitoring. While the non automatic sampling method provides the real time monitoring of harmful substances in the air. The non automatic sampling method uses the sensors to monitor the parameters, and send the data to central control center.

At present, for monitoring air pollution in wireless network the system includes the GSM, GPRS, etc. But these wireless nodes installation and maintenance are costly. That's why wireless sensor network have been rapidly developed. The wireless sensor network has many advantages application in military and industries. Many air pollution systems which monitor air pollute on due to hazardous health problems of this the government of Tiwan use various air quality monitoring systems [1]. In this they use MAC medium access control protocol for monitoring air quality in wireless sensor network. Khunarak et al. proposed E-nose electronic nose architecture for real time monitoring of indoor chemical polluted materials such as CO, NO2.

These chemical materials are highly toxic and cause respiratory failure [2]. The ARIMA prediction model is used to monitor air quality in wireless sensor network. The ARIMA model predicts the carbon dioxide level in the air in [3]. A conceptual framework for the deployment of wireless sensor network for environment monitoring of Bangalore urban city is proposed in [4]. An outdoor air pollution monitoring system which uses ZigBee networks for monitoring air pollution in ubiquitous cities was reported in [5]. This system integrates wireless sensor board and CO2 sensors which also integrates with the dust, temperature and humidity and ZigBee module. In China, Zhang Qian et al. compared the advantages of ZigBee with Wi-Fi and Bluetooth and give a wireless solution based on ZigBee technology for greenhouse monitoring [6]. Although some authors use ZigBee to monitor air pollution, its application in air pollution monitoring stay little. The exploitation of the technology of the wireless sensor network and ZigBee module, focusing on the air pollution monitoring system.

II. Pollution Parameters

The air pollution means presence of one or more contaminants for temporal duration that can become injurious to human life, vegetable, and animal. The air contaminants include smokes, gases, dusts, paper hashes, poisonous chemical products and many polluted materials.

Certain polluted materials react with each other and produces other pollutants. These pollutants called as secondary pollutants. Carbon dioxide and nitrogen dioxide produced by automobiles motors, lead to the development of ozone. Air pollution has consequences for human health. It causes respiratory harms and even fatality. It also contributes the acid rain and reduction of ozone layer.

The proposed system is able to measuring the following gases in the environment.

a. Carbon Dioxide (CO2) – Carbon Dioxide is a gas essential to life in the planet, because it is one of the most important elements evolving photosynthesis process, which converts solar into chemical energy. The concentration of CO2 has increased due mainly to massive fossil fuels burning. This increase makes plants grow rapidly. The rapid growth of unwanted plants leads to the increase use of chemicals to eliminate them.

b. Sulphur Dioxide (SO2) – Sulphur Dioxide is a pale gas, noticeable by the special odor and taste. Like CO2, it is mainly due to fossil fuels burning and to industrial processes. In high concentrations may cause respiratory harms, especially in sensitive groups. It contributes to acid rains.

c. Nitrogen Dioxide (NO2) – Nitrogen Dioxide is a brownish gas, easily noticeable for its odor, very acidic and extremely oxidant. It is produced as the result of fossil fuels burning. typically NO thrown to the atmosphere is converted in NO2 by chemical processes. In high concentrations, NO2 may lead to respiratory harms. Like SO2, it contributes to acid rains.

III. About Zigbee

The ZigBee is the small range, low power, and low data rate wireless networking technology for many wireless applications. It is present at the bottom three layers i.e. physical, data link, and network layer. This is the recently published IEEE 802.15.4 standards for personal area networks. ZigBee is embattled at radio-frequency (RF) applications that require a low data rate, extended battery life, and secure networking. ZigBee is a wireless technology developed as an open global standard to address the unique needs of low-cost, low-power, wireless sensor networks.

ZigBee network layer supports star, mesh and tree topologies. The ZigBee coordinator is responsible for initiating and maintaining the devices present in the network and other end devices directly communicate with the ZigBee coordinator.

The IEEE 802.15.4 (ZigBee) standard provides three frequency bands for operation: these are 868MHz, 916MHz, and 2.4GHz for ZigBee. 868MHz band used in only Europe and has the 20Kbps data rate of transmission and contain only one channel with BPSK modulation technique. 916MHz band is used in Americas having the 40Kbps data rate of transmission and contain 10 channels with BPSK modulation technique. 2.4GHz frequency bands used throughout the world because of ISM (Industrial, Scientific, Medical) band. It has 250Kbps data rate of transmission and 16 channels with O-QPSK modulation technique.

Transmission distance is within the range from 30 meters in an indoor non-line of sight of environment and 100 meters in line of sight environment. The range problem can be solved by using various routing algorithms at the network layer.

IV. Hardware Architecture

The proposed system is designed by integrating the following hardware modules as shown in figure.

4.1 H/W BLOCK DIAGRAM:

The following diagram shows the hardware block diagram of proposed system.

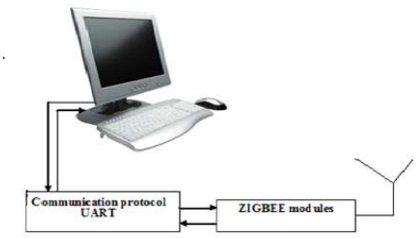


Fig. 2: System Side Block Diagram

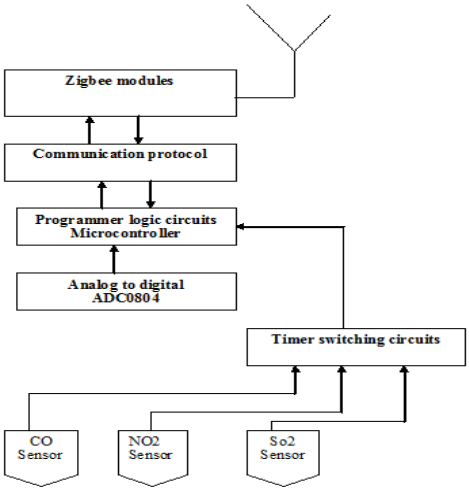


Fig. 3: Node Block Diagram

To satisfy the system's functional and nonfunctional requirements, two major building blocks are needed, namely: a Data Acquisition unit and a Pollution Monitoring Server. The data acquisition unit is designed by integrating microcontroller with a sensor array using analog ports. Data Acquisition Unit is also connected to a GPS module and ZigBee modem using RS-232 interface.

The sensor array consists of three air pollutions sensors including Carbon Dioxide (CO2), Nitrogen Dioxide (NO2), and Sulfur Dioxide (SO2). The GPS module provides the physical coordinate location of the DAQ, time and date in National Marine Electronics Association (NMEA) format [11]. NEMA format includes the complete position, velocity, and time computed by a GPS receiver where the position is given in latitude and longitude. The Pollution-Server is an off-the-shelf standard personal computer. The Pollution-Server connects to a database management system (MySQL) through a local area network (LAN).

V. System Implementation And Result

The proposed system is implemented with four transmitter module (Node1, Node 2, Node 3 and Node 4) and one receiver module. After successful implementation of proposed system the following fig. shows the hardware device snapshot and the different results taken when the system is tested successfully. The results of system testing shows the different graphs showing the various levels of pollutant materials in the environment. There are four transmitter (Node1 Node2, Node3, Node4)having similar hardware architecture. The hardware module and graphs shown below:

5.1 HARDWARE MODULE:

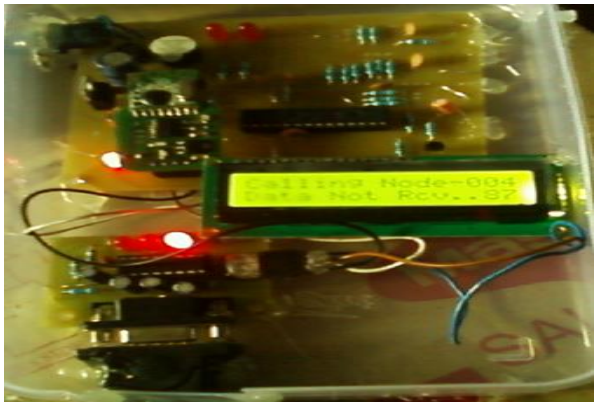


Fig. 4: Data Receiver Module

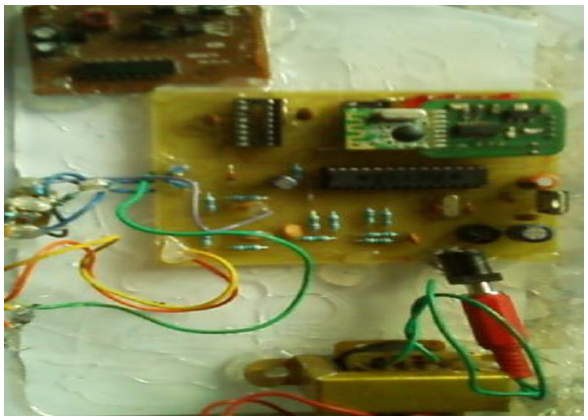


Fig. 5: Data Transmitter Module (Node 1)

5.1 RESULTS AND GRAPH:

Fig. 6: Data Monitoring Graph showing levels of NO2 for Node 1, 2, 3, and 4

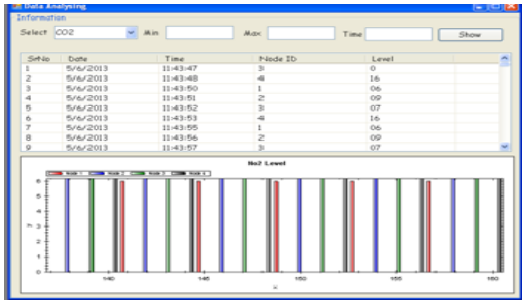


Fig. 7: Data Analyzing Graph showing pollution levels with Date and Time

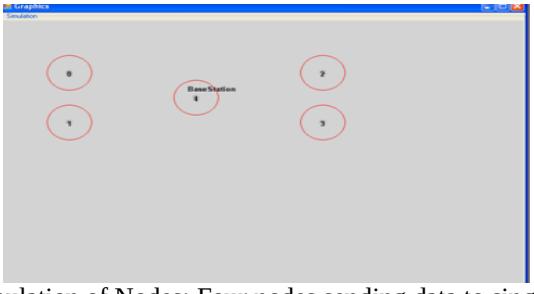


Fig. 8: Simulation of Nodes: Four nodes sending data to single receiver

VI. Conclusion

The main purpose of this paper is to provide an overview of urban air pollution monitoring application. Our work is enabled by ZigBee, pollution sensors and database sensors i.e. attached to pollution server for storing the pollutants levels for future usage by various clients for measuring pollutants in air accurately at short

intervals. The system measures air pollutant gases such as CO2, NO2, and SO2. This paper will give clear idea to move towards real time measuring in an urban area to ultimately improve quality of life on earth. Air pollution in the urban environment is a major threat to human health. As the global population is becoming more concentrated in urbanized areas, new ideas and approaches are needed to help maintain clean air that is safe for everyone to breathe. This study evaluated one such innovative approach.

VII. Acknowledgement

My sincere thanks to my honorable guide Prof. Niketa A. Chavhan and others who have contributed towards the preparation of the paper.

References

[1] Hsu-Cheng Lu , Jen-Hao Liu , Joe-Air Jiang ; Tzai- Hung Wen , Chih-Hong Sun , Jehn-Yih Juang ,"Apllication Of A Reliable MAC Protocol For The Urban Air Quality Monitoring System Based On The Wireless Sensor Network", IEEE Conference, pp. 1-6, (March 2012)

[2] Khunarak, C., Lutz, M., Kerdcharoen, T., "Wi-Fi Electronic Nose For Indoor Air Monitoring" IEEE Conference, pp. 1-4, (May 2012)

[3] Chung-Chih Lin , Ren-Guey Lee , Shi-Ping Liu , " Wireless Sensing System For Prediction Indoor Air Quality", HSIC, IEEE Conference, pp. 1-3, (May 2012)

[4] Jong-Won Kwon, Yong-Man Park, Sang-Jun Koo, Hiesik Kim, "Design of Air Pollution Monitoring System Using ZigBee Networks for Ubiquitous-City", In Proceedings Of the 2007 International Conference on Convergence Information Technology, pp.1024-1031, (2007)

[5] Zhang Qian, Yang Xiang-Long, Zhou Yi-Ming, Wang Li-Ren, Guo Xi-Shan, " A Wireless Solution For Greenhouse MonitoringAnd Control System Based On Zigbee Technology", J. Zhejiang Univ Sci A, vol. 8, pp. 1584-1587, (2007)

[6] Kavi K. Khedo, Rajiv Perseedoss and Avinash Mungur, " A Wireless Sensor Network Air Pollution Monitoring System", International Journal of Wireless and Mobile Network (IJWMN), Vol. 2,No.2, (May 2010)

[7] A. R. Al-Ali,, Imran Zualkernan, and Fadi Aloul, "A Mobile GPRS-Sensors Array for Air Pollution Monitoring", IEEE Sensor Journal, Vol. 10, No. 10, (October 2011)

[8] Jong-Won Kwon*,Yong-Man Park, Sang-Jun Koo, Hiesik Kim, "Design of Air Pollution Monitoring System using ZigBeeNetworks for Ubiquitous-City", International Conference on Convergence Information Technology, (2007)

[9] Halit Eren, Ahmed Al-Ghamdi, Jinhua Luo, "Application of ZigBee for Pollution Monitoring Caused by Automobile Exhaust Gases", IEEE Sensors Applications Symposium New Orleans, LA, USA - February 17-19(2009)

[10] Han Zhi-gang, Cui Cai-hui, "The Application of Zigbee Based Wireless Sensor Network and GIS in the Air Pollution Monitoring", International Conference on Environmental Science and Information Application Technology, (2009)

[11] W. Chung and C. H. Yang, "Remote monitoring system with wireless sensors module for room environment," Sens. Actuators B, vol. 113 no. 1, pp. 35– 42,(2009)

[12] M. Predko, Programming and Customizing PICmicro Microcontrollers, Mc Graw Hill, (2002)

[13] P.N. Borza, C. Gerigan, P. Ogruţan, Ghe. Toacşe, Microcontrollers – Applications (in Romanian), Ed. Tehnică, Bucuresti, Romania, (2000)

[14] O. Hyncica, The ZigBee Experience.

[15] Y W Zhu,The Design of Wireless Sensor Network System Based on ZigBee Technology for Greenhouse, 2006 IOP Publishing Ltd, pp1195~1199

[16] ZigBee Alliance, ZigBee-Specification National Marine Electronics Association Data. [Online]. Available: http://www.gpsinformation.org/dale/nmea.htm

Authors

Darshana Tambe received the B.E. degree from K.D.K College of Engineering, Nagpur, State-Maharashtra, India. She is pursuing Master of Engineering (M.E.) in Wireless Communication and Computing from G. H. Raisoni College of Engineering, Nagpur. Maharashtra, India. Her research area includes Wireless network security, Wireless sensor network.

2.

3.

4

Niketa A. Chavhan received the Master of Engineering (M.E.) in Wireless Communication and Computing from G. H. Raisoni College of Engineering, Nagpur, Maharashtra, India. She is working as Assistant Professor in G. H. Raisoni College of Engineering, Nagpur. Her research area includes Ad-hoc Wireless networks, Wireless sensor networks and Mobile Technology.