AVOIDING PRIVACY LEAKAGE TO CLOUD SERVER WITH PRIVACY PRESERVING TO SQL QUERIES

[1]Asma Mohammed, [2]Dr. R. Jegadeesan [3]G.RanjithKumar, [4]M.Ragini, [5]N.Sruthi, [6]N.Rasagna

[1,4,5,6] B.Tech Final Year Students, [2,3]Associate Professor Dept of CSE

[1,2,3,4,5,6]Jyothishmathi Institute of Technology & Science, Karimnagar, India.

Abstract: Recent days, outsourcing database to cloud is getting popular as cloud provide services and applications which are in low cost and available even foe small scale industries. There are different schemes to provide appropriate functionality for the queries over the outsourced database. Though, presence of such schemes provide sufficient functionality, the privacy may get leaked to the cloud which is providing service due to the frequent query search over the outsourced database as it is in total control of cloud server. When we consider numerical range queries, they cannot have secure schemes which can leak there statistical properties as well as access patterns which is an important practical challenge. To avoid this privacy leakage, this paper provide a multi cloud scheme by dividing the knowledge into these clouds, which is efficient enough to provide security to outsourced database and to numeric range queries over database.

Index Terms – database, cloud computing, range query.

1.INTRODUCTION

Cloud computing involves services over the internet. It has services like IaaS, PaaS, SaaS[1].One can outsource the data to the cloud and can view whenever the user desired. computing saves the cost, provides security, flexibility etc[2]. Cloud has a privacy issue that ,the service provided by the cloud is assumed to be semi-trust i.e., the cloud is honest but curious .As the cloud is assumed to be semi-trust, the data of the of the data owners is at risk. So, a technique called encryption should be done to the data before outsourcing to the cloud.

An overview for outsourced database is explained in cryptDB[3]. Consider any cloud client like software company. The private data has to be send to the cloud and can be accessed using DDL, DML, DQL, DCL commands[5]. As we have assumed that the cloud is semi-trust, there is a chance for the cloud to get any private information for the advantage of the company. In some worst situations, there is a chance for the cloud to leak the private information to the opponents for profit which is illegal.

The privacy can be stored irrespective of the cloud. The confidential data is divided into two parts and distributed into two non-colluding clouds. The implementation of divide and conquer method can know any confidential information from single part of knowledge and each cloud will have idea about it's own part. A secure two cloud database architecture is introduced where the clouds are non-colluding and both the clouds have idea about it's own part. By taking this architecture into consideration we further recommend a series if interaction protocols which are the communication scenario between individual agents in multi-agent[10] systems. For a client to conduct numeric related query over enciphering data from faraway cloud servers. It involve query statements such as greater than ,less than, between etc.

1.1 MOTIVATION:

Providing and preserving security is a key factor in cloud. In modern days, everyone is concentrating on privacy issues as it is the major risk. Even though there are many schemes regarding the privacy, there are some chances to leak the private data. Many organizations and enterprises are facing these problems. If the private data of any organization or enterprise got leaked, then it has to face many problems. The goal is to provide security for the data which is stored in cloud.

2.RELATED WORK

Due to the increasing popularity to retrieve data with similar(not only same indexes) the fuzzy searchable encryption [11] introduced in many literatures for cloud computing. These search techniques allows small- scaled distinction in character or numeric level in search keywords. More importantly for numeric keywords , the predicate of query can get numeric records with in range. Some existing range query schemes due to large storage overhead to maintain the encrypted data is not suitable for practical.

Subsequently, to provide numeric related range query in database the scheme like Order Preserving Encryption (OPE) [4] is introduced . The OPE preserves the order of values in encryption field , while hiding the actual values OPE developed to increase both efficiency and security . In Ideal security OPE scheme [4] , an adversary even having the access privilege to a set of cipher text. It can not learn the knowledge of the data with non- negligible advantage. In order preserving encryption scheme [4],[6] , though it is achieved the security boundary it does not totally satisfy the privacy requirement as it exposes the order of data which OPE store inherently, this may be utilized to reveal the some amount of sensitive data.

In Security and privacy enhancing multicloud architecture [10] explained a architecture o2f cloud which or protect the sensitive information of outsourced databases and services . This mainly concentrated on the four knowledge partition patterns (1) Application Replication (2) Division of application logic (3) Fragmentation of application logic (4) Fragmentation of application data . Here , The knowledge is divided in to two fragments and stored in two different clouds which is non-colluding an each cloud knows only its respected fragmented knowledge the cloud can not get any private information in such type of multicloud architecture.

The security and privacy enhancing multi-cloud architecture [10], does not provided a brief scheme or realization for outsourced structured database.

This paper, introduced a multi-cloud prototype in which two clouds are non-colluding , in which the structured database is outsourced and also recommend a series of interaction protocols to conduct queries of outsourced database. The scheme not only provide privacy preservation but also privacy to logical relationship among data contents , i.e., order of data , stastical properties and pattern of query.

3.SYSTEM ARCHITECTURES AND PRIVACY REQUIREMNTS

In existing system, There is a cloud which stores the database which is in encrypted format though this database is encrypted there is a chance of leakage of data to the cloud server with is sensible to the client through frequent quires over outsourced database . Because of privacy related issues the cloud service provider is supposed to be semi-trust. Putting sensitive data into the cloud is a disapproving concern. So the sensitive data need to be encrypted before outsourcing the data. The cloud owner in the need of benefits will forward the sensitive data to others which are competitors. In this, the frequently asked queries of the client will unavoidably and slowly reveal some amount of private information. Some specific type cryptology like Order Preserving Encryption (OPE) [4],[6],[7] reveal the sensitive information of the clients by maintaining the order of the cipher texts. As a solution to the above problem the recommended system has two clouds cloud A and cloud B. These two clouds are used to perform different tasks , such as cloud A provides storage service and it stores the encrypted databases. While cloud B only performs the computation tasks .These computation tasks are to solve the numerical records that satisfy the client query request by using its security key. Key knowledge of application can be divided into two parts, where one part is only known to one cloud. Hence the single part of knowledge that cannot release the privacy of data and queries.

3.1. System Architecture

Our system architecture includes two non-colluding clouds and a database owner. Here database owner acts as a client in the perspective of cloud and two non-colluding clouds acts as servers. They also provide storage and related computation for query processing. The figure shows the detailed architecture of our project.

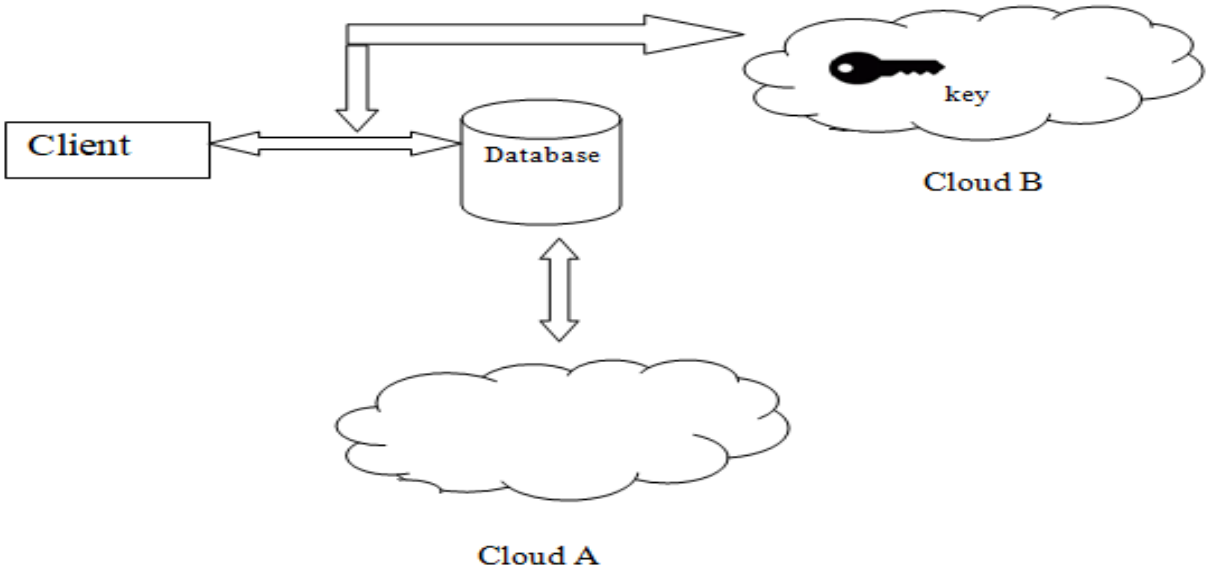

Fig.1. Two-Cloud Database Architecture

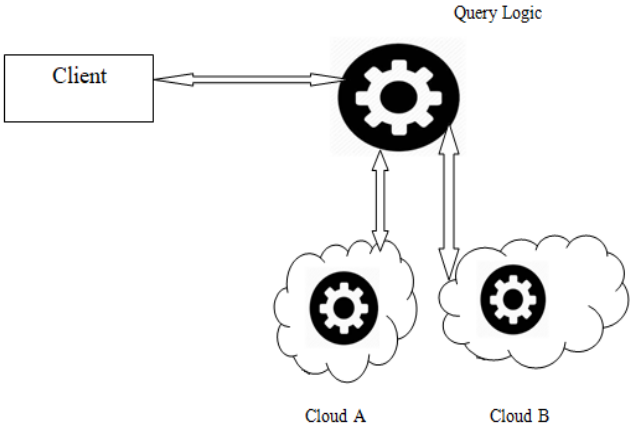

Fig.2. Knowledge Partition Prototype

In our scheme, the related knowledge of the database and queries is divided into two parts. The database is stored in one cloud and the related computation work is processed in another cloud. With this division of knowledge, every cloud will have its own data and knowing of only one part of the data is not enough to obtain any useful privacy information. This architecture of our recommended system increases some complexity but at the same time it provides security. There are many encryption schemes on data though they are efficient somehow they leak the data and fail to provide privacy preservation.

3.2 Privacy Requirements

In our system data contents, statistical properties and query patterns are the main privacy concerns.

Data contents: The privacy preservation data contents in the databases are the major concern for privacy threats. It includes the privacy of definition and description of each column in the table of the stored database and also each record value in the table. To ensure privacy for this data tables some related works have used cryptDB in which column names are encrypted and mean while the columns names are encrypted with other techniques like Order Preserving Encryption. However using encryption alone cannot provide complete security to data contents with the development of data analysis, one can extract the features from data and queries and the classification technique can help to understand the definition of columns.

Statistical properties: Although statistical properties can disclose the private information of data contents, these properties themselves are already sensitive and private for client. One of the widely used encryption scheme to construct a secure database is Order Preserving Encryption(OPE),with the uses range queries, it may leak the privacy information of encrypted field by query processing many query request over encrypted secure database.

Query pattern: The query pattern contains privacy information such as client purpose of the query and they also reveal some statistical properties as we discussed above.

To preserve privacy for above threats the outsourced database which provide numeric-related queries have to avert the following information from being obtained by semi-trusted cloud.

- For data contents: The column names and values need to be protected. For statistical properties: The statistical properties such as order of data and their probability distributions are to be protected against semi-trust cloud as some properties include '>','<','=',& 'BETWEEN' etc.
- For query pattern: The queries over secure database are need to be kept private against semi-trust clouds and also against clouds unauthorized parties such that even though it process many queries it should not reveal private information.

4.PRELIMINARIES AND SOME DEFINATIONS

4.1 Paillier Cryptographic Algorithm

Key generation:

Step 1: Take two prime numbersk and l independently and randomly suchthat$gcd(kl,(k-1)(l-1))=1$.

Step 2: Evaluate m=kl and λ=lcm(k-1, l-1).

Step 3: choose random integer 'p' where p ϵ $Z^*_{n^2}$.

Step 4: Evaluate μ=$(F(\ p^{\lambda} mod\ m^2))^{-1} mod\ m$ where function F is defined as

$F(x) = \frac{x-1}{m}$.

- The encryption (public) key is *(m, p).*
- The decryption (private) key is (λ , μ).

Encryption:

Step 1: Let b be a message to be encrypted where $0 \leq b < m$.

Step 2: Choose a number 'q' where $0 < q < m$ and $q \in Z^*_{n^2}$ (i.e., ensure $gcd\ (q\ ,\ m)=1$).

Step 3: Compute cipher text as: $c = p^b . q^m mod m^2$.

Decryption:

Step 1: Let c be the cipher text which need to be decrypt, where $c \in Z^*_{n^2}$.

Step 2: Compute the plain text message from cipher text message as $b=F(c^{\lambda} mod m^2).\ \mu\ mod\ m$.

4.2 Numeric – Related SQL Queries

The SQL (structured query Language) is a programming Language, which is used to perform operations on the relational database and to manage data in the database.A query operation is a request data with a statement to explain the desired data. The data which is required can be a single column or more than one column of one or more tables in a outsourced database and it aggregate the results from original data, for example sum , average , count , of the data. To retrieve the designed data the query must be describe the requirement as the numeric-related SQL queries in the rest of paper.The introduced two-cloud architecture , propose a series of interaction protocols between the client and the clouds , which can realize numeric- related SQL queries and obtain the privacy requirements .Apart from retrieving data with query operation , there are other SQL operations like Update, Insert which can alter the data. The privacy preservation issue for this data is revolve with other approaches which are existing like ORAM (Oblivious RAM) [12].This paper concern only on the privacy preservation to query operation.

5. METHODOLOGY

This section provide a complete interaction protocols to realize range query with privacy preservation on database which is outsourced to a cloud. Here the schemes includes the two clouds which are non-colluding and the knowledge of application is divided and that divided knowledge is shared between two clouds. The cloud A usually contain the outsourced database i.e., storage and the cloud B will do the computation task.

The two cloud architecture provide privacy preservation to numeric related data. The client can retrieve the data from queries when the query includes the operations like '>', '<', and 'BETWEEN' for columns.

For example,

SELECT * FROM Table WHERE H_i> c.

In the above query the client wants to retrieve the items from the table, whose column H_i should be greater than a constant 'c'. In this scheme, it is resolved by taking the sign of each value of (H_i(k) - c), where 'k' traverses all rows of the whole table if the result of (Hi(k) - c) is greater than '0', the item set which are relevant to the result satisfies the query predicate and if the operator is reversed i.e., if the condition become "H_i(k) < c" then the corresponding computation will also be get reversed as (c-H_i(k)) and the remaining phase are similar to that of the case '>'.

For the 'BETWEEN c and d' predicate the result is retrieved through the intersection of H_i> c and H_i< d.

The recommended scheme is collection of table creation and query protocol. The query protocol har four parts according to its operation.

1) Query request
2) Sending the item
3) Sending the index
4) Query response

Table creation: After the renting cloud service. The client will outsource the database to cloud before outsourcing the database application the client need to provide protection in the following way.

- The column name of the table H_i should be encrypted [i.e., denoted by E(H_i)] and the key 'k' should be securely kept by the client.
- For each item(The row in a table), its values also need to be encrypted. This paper only concentrate on the numeric related data and it uses Palliers cryptosystem which make use of public key for each numeric related value "y".
  Y=E(y , PK)
- After encryption of database it needs to be uploaded into the cloud A along with public key and the secret or private key will be sent to cloud B securely.

Query Protocol:

Here we explain query protocol by implementing the recommended procedure for operator ">".

1) Query Request: After outsourcing the database, whenever the clients wants to retrieve data from the outsourced database they need to generate the request in the form of SQL query.

For example,

SELECT * FROM table WHERE H_i> c;

After generating the plain text query it need to be converted to cipher text in order to preserve the privacy.

i) The encryption of the column name: The client need to be encrypt the column name as E(H_i>) with key 'K'.

ii) The encryption of range boundary value: After encrypting the column name the client also need to encrypt the range boundary value with the public key PK in Pallier cryptosystem. After encrypting the boundary value it is denoted as 'C'.

iii) The Token generation: After completion of above process the client generate the token to analyze query request. The token is in the form of *Sign(TN || CN ||N||T),*where

TN = Token serial number
CN = Number of the column involved
N = Total number of items in the table
T = The current timestamp

All these are signed by the client's secrete key SK..

iv)Sending the query request: At last, the client send the following encrypted query to the cloud A with the signed token. Here the CN is '1'.

SELECT * FROM table WHERE E(H_i>) > C.

2) Sending the Item: After getting the query request the Cloud A process the query and retrive the column name E(H_i) and process following phases:

Phase 1:

For each item Y_j=H_{ij} in the column ,two random positive numbers are selected by the cloud A as r_j and ⍰$_j$where $0 \le$ ⍰$_j <$ r_j and computes ,

$$Y_j' = (\frac{Y_j}{C})^{r_j}\ E(-⍰_j, PK)$$

Phase 2:

After phase 1, the cloud A shuffles the retrieved array randomly to generate new array of items and the cloud A keep the mapping between the old array and the newly formed array securely.

Finally, the cloud A removes the Column name E(H_i) from newly formed array and send to cloud B along with the token .

3)Sending the Index: After receiving the array and token from cloud A the cloud B checks the genuineness token to make sure it has not expired and has not used in specific time. IT also checks the column number and item number consist with the values in token and if the request is authorized then cloud B decrypts each as shown.

$$y'_{j'} = D(Y'_{j'}, SK)$$

The each decrypted item is inserted into a another new array in the cloud B if it is greater than 0 .To preserve the privacy the cloud B add some dummy indexes to random positions of the newly formed array and return this array to the Cloud A.

4) Query Response:

For each received item from cloud B the cloud A loose up the index information and gets corresponding index in the original column according to the mapped index. Cloud A sends the corresponding rows in the table, asquery response to the client. After client receiving the response, the client decrypt the items with private key and removes dummy items which do not satisfy the query predicate.

Different schemes for Operators '<','BETWEEN','=' :

1)Operator '<': The query processing for '<' operator is similar to '>' the difference lies in the phase 1 that is here in order to implement subtraction (c-H_i(k)) the corresponding operation in encryption is,

$$Y'_j = (\frac{C}{Y_j})^{r_j}\ E(-\square_j, PK)$$

2) For Operator 'BETWEEN' and '=' :

The equivalent to an 'AND' logic for operator 'BETWEEN' in the query is (H_i> c) ^ (H_i< b)

The operator '=' can be treated as special case of 'BETWEEN' and the predicate 'H_i=c' is translated as 'Hi BETWEEN c-1 AND c+1" and the equivalent AND logic as follows

(H_i> c-1) ^ (H_i< c+1).

7. FUTURE ENHANCEMENT

In therecommended system, we will contemplate to further raise the security level while protecting the feasibility and we will expand our future scope to bear more operations like SUM\AVG and this our system is limited to numerical queries only, for a future scope it can be enhanced to support other type of queries .

8.CONCLUSION

Here, we furnished a two-cloud plan with a chain or sequence of interaction protocols for outsourced database facility, which safeguard the privacy preservation of data contents, statistical property and query pattern. Parallelly, with the help of range queries it not only secures the privacy of static data but also label standard privacy outflow in statistical properties or after many query processes.Security (or) certainty examination reveal that our strategy can encounter the privacy preservation needs. The further performance appraisal reveal that our recommended system is efficient.

REFERENCES

[1] M Armbtust, A Fox , R.GRIFFITH, A.D. Joseph et al ,"A view of cloud computing" ,Communications of the ACM ,2010.

[2] Cong Wang, Student Member, IEEE, Qian Wang, Student Member, IEEE, Kui Ren, Senior Member, IEEE, Ning Cao, and Wenjing Lou, Senior Member, IEEE, "Toward Secure and Dependable Storage Services in Cloud Computing", IEEE, APRIL-JUNE 2012.

[3] R.APopa,C.Redfield,N.Zeldovich and H.Balaakrishnan, "CryptDB: protecting confidentiality with encrypted query processing", ACM,2011.

[4] Ada Popa, Frank H. Li ,NickolaiZeldovich, "Ideal-Security Protocol for Order-Preserving Encoding", IEEE 2013.

[5] Valentina Ciriani, Sabrina De Capitani di Vimercati, Sara Foresti, Sushil Jajodia, Stefano Paraboschi, and PierangelaSamarati ," Keep a Few: Outsourcing Data While Maintaining Confidentiality", Spirnger 2009.

[6] Hasan Kadhem,Toshiyuki Amagasa,Hiroyuki Kitagawa, **"**Optimization Techniques for Range Queries in the Multivalued-partial Order Preserving Encryption Scheme**",** Springer 2010.

[7] R.Agarwal,J.Keirnan,R.Srikant,Y.Xu**,"**Order preserving encryption for numeric data",ACM 2004.

[8] E.Shi,J.Bethencourt,T.H.Chan,D.Song and A.Perrig, "Multi-dimensional range query over encrypted data",IEEE 2007.

[9] P. Paillier, "Public-key cryptosystems based on composite degree residuosityclasses",May ACM 2000.

[10]D. Ardagna, "Cloud and Multi-cloud Computing: Current Challenges and Future Applications," 2015 IEEE/ACM 7thInternational Workshop on Principles of Engineering Service-Oriented and Cloud Systems, *Florence, 2015, pp. 1-2.doi:* 10.1109/PESOS.2015.8.

[11] F.Hao, J. Daugman and P.Zielinski, "A fast search algorithm for a large fuzzy database",IEEE ,2008.

[12]E.Stefanov and E. Shi ,"Multi-cloud oblivious storage," ACM ,2013.

[13].Jegadeesan,R.,Sankar Ram M.Naveen Kumar JAN 2013 "Less Cost Any Routing With Energy Cost Optimization" International Journal of Advanced Research in Computer Networking,Wireless and Mobile Communications.Volume-No.1: Page no: Issue-No.1 Impact Factor = 1.5

[14]. Jegadeesan,R.,Sankar Ram, R.Janakiraman September-October 2013
"A Recent Approach to Organise Structured Data in Mobile Environment" R.Jegadeesan et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (6) ,Page No. 848-852 ISSN: 0975-9646 Impact Factor:2.93

[15]. Jegadeesan,R., Sankar Ram October -2013 "ENROUTING TECHNICS USING DYNAMIC WIRELESS NETWORKS" International Journal of Asia Pacific Journal of Research Ph.D Research Scholar 1, Supervisor2, VOL -3 Page No: Print-ISSN-2320-5504 impact factor 0.433

[16]. Jegadeesan,R., Sankar Ram, M.S.Tharani (September-October, 2013)
"Enhancing File Security by Integrating Steganography Technique in Linux Kernel" Global journal of Engineering,Design & Technology G.J. E.D.T., Vol. 2(5): Page No:9-14 ISSN: 2319 – 7293

[17]. Ramesh,R., Vinoth Kumar,R., and Jegadeesan,R., January 2014 "NTH THIRD PARTY AUDITING FOR DATA INTEGRITY IN CLOUD" Asia Pacific Journal of Research Vol: I Issue XIII, ISSN: 2320-5504, E-ISSN-2347-4793 Vol: I Issue XIII, Page No: Impact Factor:0.433

[18]. Vijayalakshmi, Balika J Chelliah and Jegadeesan,R., February-2014 "SUODY-Preserving Privacy in Sharing Data with Multi-Vendor for Dynamic Groups" Global journal of Engineering,Design & Technology. G.J. E.D.T.,Vol.3(1):43-47 (January-February, 2014) ISSN: 2319 –7293

[19]. Jegadeesan,R.,SankarRam,T.Karpagam March-2014 "Defending wireless network using Randomized Routing process" International Journal of Emerging Research in management and Technology

[20].Jegadeesan,R.,T.Karpagam, Dr.N.Sankar Ram , "Defending Wireless Network using Randomized Routing Process" International journal of Emerging Research in management and Technology ISSN: 2278-9359 (Volume-3, Issue-3) . March 2014

[21]. Jegadeesan,R., Sankar Ram "Defending Wireless Sensor Network using Randomized Routing "International Journal of Advanced Research in Computer Science and Software Engineering Volume 5, Issue 9, September 2015 ISSN: 2277 128X Page | 934-938

[22]. Jegadeesan,R., Sankar Ram,N. "Energy-Efficient Wireless Network Communication with Priority Packet Based QoS Scheduling", Asian Journal of Information Technology(AJIT) 15(8): 1396-1404,2016 ISSN: 1682-3915,Medwell Journal,2016 (Annexure-I updated Journal 2016)
[23]. Jegadeesan,R.,Sankar Ram,N. "Energy Consumption Power Aware Data Delivery in Wireless Network", Circuits and Systems, Scientific Research Publisher,2016 (Annexure-I updated Journal 2016)
[24]. Jegadeesan,R., Sankar Ram , and J.Abirmi "Implementing Online Driving License Renewal by Integration of Web Orchestration and Web Choreogrphy" International journal of Advanced Research trends in Engineering and Technology (IJARTET) ISSN:2394-3785 (Volume-5, Issue-1, January 2018
[25]. Pooja,S., Jegadeesan,R., Pavithra,S., and Mounikasri,A., "Identification of Fake Channel Characteristics using Auxiliary Receiver in Wireless Trnsmission" International journal for Scientific Research and Development (IJSRD) ISSN (Online):2321-0613 (Volume-6, Issue-1, Page No. 607-613, April 2018
[26]. Sangeetha,R., Jegadeesan,R., Ramya,P., and Vennila.,G "Health Monitoring System Using Internet of Things" International journal of Engineering Research and Advanced Technology (IJERAT) ISSN :2454-6135 (Volume-4, Issue-3, Page No. 607-613, March 2018.