A REVIEW STUDY ON DISTRIBUTED DATABASE COMMUNICATION

[1]Amneet Kaur, [2]Mrs. Meenakshi Bansal
[1]M.tech Student, [2]Assistant Professor
Computer Engineering Department,Yadavindra College of Engineering,
Punjabi University, Patiala

ABSTRACT: Distributed database plays a vital role in day to day life because in the present era, business environment is increasing at very fast rate so our basic desire is to get reliable information from any source. Since our database is distributed, means data is located at different geographical locations and finally helps to easily access our valuable & precious data. We propose a architecture that integrates cloud database services with data integrity and the possibility of executing concurrent operations on encrypted data. It is the solution supporting geographically distributed clients to connect directly to an encrypted cloud database, and to execute the concurrent and the independent operations including those modifying the database structure. Distributed database is the emerging technique which focuses on concurrency control and security issues under this distributed database. The NTRU being fast and secure hashing algorithm will provide more security to the system, in terms of throughput and processing speed.

1. INTRODUCTION

Distributed database plays a important role in our daily life, because now a day's business environment is increasing at very fast rate, therefore due to this our basic aim is that the information we receive from any source should be accurate and reliable as per our need, because if the information is not reliable than it is useless for the user. As our database is distributed it itself defines that the data is located at the different geographical locations and also helps us to access the reliable and valuable data. In case of normal database we will face the problem of failure at one point means overall failure but in case of distributed database such kind of problem will not occur. This is because the data is geographically located at more than one location. If there are the chances of failure at one location we can access the data from another location.

1.1 Distributed Database System

A distributed database is defined as a database in which the storage devices are not all connected to a common processing unit such as CPU, controlled by a distributed database management system. It may be stored in multiple computers, located in the same physical location or may be dispersed over the network of interconnected computers. The distributed database system consists of the loosely-coupled sites that will share no physical components. The system administrators can distribute the collections of data across the multiple physical locations. A distributed database can reside on the network servers on Internet or on the other company networks because they store the data across the multiple computers, the distributed databases can improve performance at end-user worksites.

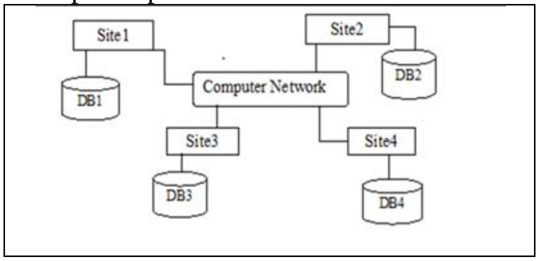

Fig -1.1: Distributed database systems [10]

In figure1.1 of distributed database their present three sites with their respective databases, all of these are connected through the means of computer network.

Distributed database has many benefits due to which it is widely used in business organization because main factor which it includes is performance. The two processes which ensure that the distributed databases remain up-to-date and current are replication and duplication.

1. Replication: it makes the use of specialized software that looks for changes in the distributive database. A replication process makes all databases look the same once the changes are identified. This process may be a complex and time-consuming depending on number or the size of the distributed databases. It requires a lot of time and the computer resources.

2. Duplication: it has less complexity. It recognizes one database as a master and then duplicates that database. The duplication process is done at the set time after hours. This have to ensure that each distributed location have the same data. In the duplication process, users can change only the master database. This ensures that local data will not be overwritten. The database systems that run on each site may have a substantial degree of the mutual independence. As the distributed networks become more popular, the need for improvement in the distributed database management systems becomes even more important.

1.2 Distributed Database Design

Distributed database design plays the important role, because not only of its performance but also because of its reliability and the efficient way of reducing the overall cost and hence it will act as a saving solution to our problem, as systematic design can give solution to many problems. The design phase includes two approaches[1]

- View design - defining the interfaces for end users.
- Conceptual design - is the process by which the enterprise is examined to determine entity types and relationships among these entities. One can possibly divide this process into to related activity group.
- Distributions design - design the local conceptual schemas by distributing the entities over the sites of the distributed system. The distribution design activity consists of two steps: Fragmentation and Allocation.
- Physical design - is the process, which maps the local conceptual schemas to the physical storage devices available at the corresponding sites,
- Observation and monitoring - the results is some form of feedback, which may result in backing up to one of the earlier steps in the design.

Top down approach: this approach generally follows hierarchical pattern i.e. information flows in hierarchy that means first of all defines the generally used concepts followed by the framework that we are going to use and then the detail.

Bottom down approach: this is generally the reverse of the top down approach, instead of defining the framework, this approach focus on details first then later on deal with the framework. There are present some additional factors that are considered under this concept along with these top down and bottom down approach [1]

Data fragmentation: it means that preparing the fragments of any particular data or we can say data is divided into different parts and each part is processed separately, this technique will improve the overall performance with appropriate speed. Fragmentation can be done by different ways i.e. horizontally (division is done horizontally with the selection of particular row), vertically (division is done vertically mainly by selecting the relevant column), mixed (done by using some operators that can be either union or join relevant to our problem). Data replication: In distributed database as database is distributed at many locations therefore fragments of the data should be there on all the locations, therefore with the help of data replication it become easy in deciding that which particular fragment or part of data should be replicated so that it can be replicated at different locations

Data allocation: Allocation as the name itself specifies, allocating or giving the data to all the sites where our data is distributed. Therefore the fragment will be given or distributed among multiple sites and then it will be stored, so that it can be used at the time of need.

1.3 DISTRIBUTED DATABASE SECURITY & CONCURRENCY CONTROL

As the technology is growing day by day the problems related to the security of the distributed database system are also increasing due to which the security becomes an important issue. The distributed database has all of the security concerns of a single-site database plus several additional problem areas. We begin our investigation with a review of the security elements common to all database systems and those issues specific to distributed systems. Some commonly occurred security problems unique to Distributed Database Management Systems are:

- Centralized or Decentralized Authorization.
- In developing a distributed database, one of the first questions to answer is where to grant system access. Bell and Grisom outline two strategies: Users are granted system access at their home site/ remote site.
- Users are granted system access at their home site-success of this strategy depends on reliable communication between the different sites. Since many different sites can grant access, the probability of unauthorized access increases.
- Users are granted system access at the remote site-This strategy, while perhaps more secure, has several disadvantages.
- Probably the most glaring is the additional processing overhead required, particularly if the given operation requires the participation of several sites.
- Preservation of integrity is much more difficult in a heterogeneous distributed database than in a homogeneous one.
- The homogeneous distributed database has strong central control and has identical DBMS schema
- If the nodes in the distributed network are heterogeneous several problems can arise that will threaten the integrity of the distributed data.
- These inconsistencies can cause problems, particularly with complex queries that rely on more than one database
- Conflict resolution depends on the level of central control.

1.4 REQUIREMENTS OF SECURE DATABASE SYSTEM

A secure database must satisfy the following requirements

1. It must have physical integrity.
2. It must have logical integrity.
3. It must be available when needed.
4. The system must have an audit system.
5. It must have elemental integrity (accurate data).
6. Access must be controlled to some degree depending on the sensitivity of the data.
7. A system must be in place to authenticate the users of the system.
8. Sensitive data must be protected from inference.
9. Password authentication for users and roles
10. Some types of external authentication for users and roles including:
11. Login packet encryption for client-to-server and server-to-server connections

The key goal of these requirements is to ensure that data stored in the DBMS is protected from unauthorized observation or inference, unauthorized modification, and from inaccurate updates. This can be accomplished by using access controls, concurrency controls, and inference reduction strategies.

Concurrency: Concurrency means that the users access the information from the database concurrently but sometimes there arise the problem when two or more users try to access the same problem at the same time.[3] Therefore concurrency should be controlled that can be done by adopting the proper serialization concept while executing more than one transaction at the same time its necessary to take proper measures to control the concurrency. This concurrency should be controlled, as in distributed database their present different servers that were located at various locations.

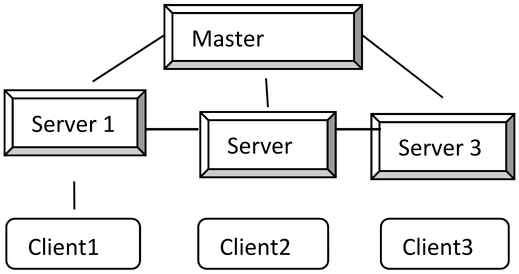

Fig 1.2: Updations in distributed database

Suppose if any client update something on the servers than the information will go to the masters and then masters will be responsible to update the contents to the other related servers. This way the whole work proceeds. In this case suppose we have four clients and three servers and client one update something in one of the three servers then the change should be updated in other three servers. If the updated content is not reflected in on other three servers their may arise the conflict.

1.5 Cloud Database

Cloud database management system (CDBMS) is defined as the distributed database that delivers a query service across multiple distributed database nodes located in the multiple geographically-distributed data centers, both corporate with the cloud data centers. Database accessible to the clients from the cloud will be delivered to the users on demand via Internet from a cloud database provider's servers.

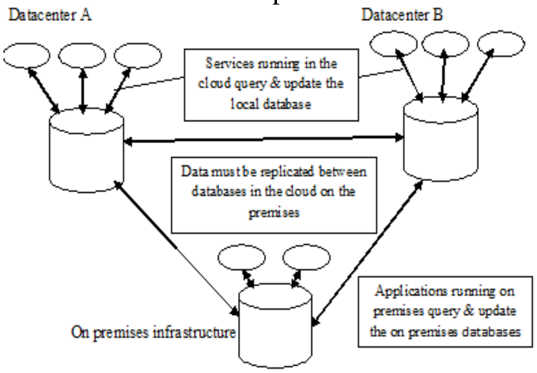
Datacenter A
Datacenter B
Services running in the cloud query & update the local database
Data must be replicated between databases in the cloud on the premises
Applications running on premises query & update the on premises databases
On premises infrastructure

Fig-1.3: Cloud Database

The cloud databases can use cloud computing to achieve optimized scaling, high availability and effective resource allocation. While cloud database can be a traditional database such as a MySQL or SQL Server database that has been adopted for the cloud use. Cloud databases can offer significant advantages over their traditional counterparts, includes increased accessibility, automatic failover and the fast automated recovery from the failures, minimal investment and potentially have the better performance.

1.6 Advantages of Distributed Database [9]

1) In the distributed database, data can be stored in the different systems like personal computers, servers, mainframes, etc.

2) A user doesn’t know where the data is located physically. The Database presents the data to the user as if it were located locally.

3) Database can be accessed over the different networks.

4) Data can be joined and updated from the different tables which are located on the different machines.

5) Even if the system fails the integrity of the distributed database is maintained.

6) A distributed database is secure.

1.7 Application of Distributed Database

1. Banking: For customer information, accounts, and loans, and banking transactions.

2. Airlines: For reservations and schedule information. Airlines were among the first to use databases in a geographically distributed manner - terminals situated around the world accessed the central database system through phone lines and other data networks.

3. Universities: For student information, course registrations, and grades.

4. Credit card transactions: For purchases on credit cards and generation of monthly statements.

5. Telecommunication: For keeping records of calls made, generating monthly bills, maintaining balances on prepaid calling cards, and storing information about the communication networks.

6. Finance: For storing information about holdings, sales, and purchases of financial instruments such as stocks and bonds.

7. Sales: For customer, product, and purchase information.

8. Manufacturing: For management of supply chain and for tracking production of items in factories, inventories of items in warehouses / stores, and orders for items.

9. Human resources: For information about employees, salaries, payroll taxes and benefits, and for generation of paychecks.

2. LITERATURE SURVEY

Bhullar *et al.*[2014] [4] Distributed database is the emerging technique that plays an important role, in day to day life, so we focus on concurrency control and security issues under this distributed database. In this paper we are going to analyze the NTRU algorithm, based Encryption decryption technique for security to Distributed Database System.

Ferretti *et al.*[2014] [6] revealed that placing critical data in the hands of a cloud provider should come with guarantee of security and availability for data at rest, and in motion. Some alternatives exist for the storage services, while the data confidentiality solutions for the database as a service paradigm are still immature. It describes the architecture that integrates cloud database services with data confidentiality and the possibility of executing the concurrent operations on encrypted data.

Ferretti *et al.*[2014] [8] revealed that the success of the cloud database paradigm is strictly related to strong guarantees in terms of service availability, scalability and security, but also of data confidentiality. Any cloud provider assures the security and availability of its platform, while the implementation of scalable solutions to guarantee confidentiality of the information stored in cloud databases is an open problem left to the tenant. Existing solutions address some preliminary issues through SQL operations on encrypted data. We propose the first complete architecture that combines data encryption, key management, authentication and authorization solutions, and that addresses the issues related to typical threat scenarios for cloud database services.

Gill *et al.*[2014] [2] In these days, the security is essential for all the applications on the network. For providing the security to many applications on network, numbers of mechanisms are used. But there is no implementation of any mechanism for security on the N-tier architecture. The NTRU algorithm is concluded as a best and fast algorithm for providing security on the clouds.

Hababeh [2010] [5] says that customizing network sites have become an increasingly important issue in distributed database systems. This will improve the network system performance by reducing the number of communications required for query processing in terms of retrieval and update transactions. This paper presents an intelligent clustering method for distributed database system that provides a structure for organizing large number of network sites into a set of useful clusters to minimize transactions processing communications. It has been designed to divide the database network sites into a set of disjoint clusters based on a high performance clustering technique. This can reduce the amount of redundant data to be accessed and transferred among different sites, definitely increase the transaction performance, database system response time, and result in better distributed network decision support.

Mekala *et al.*[2014] [11] Cloud computing is one of the most increasing one with the increase number of cloud users. In today's environment every user wants to access their data at any time and at anywhere. In an organization they store the data only on their computers, if they want their data during the roaming situation means it is not possible one to carry the data at every time, this is a difficult factors for an organization. The Cloud computing can address this problem by providing data storage mechanism to access the data at anywhere. This is the storage device used to access their data at any where through networks which is called cloud provider.

Yashpal Mote *et al*.[2012] [13] in their paper says about data encryption algorithm, this algorithm play important role in encrypting and decrypting the data there are present various types of the algorithm that are AES, DES, Triple DES, NTRU, and each of these algorithm have their specific role in day to day life . The two main characteristics that identify and differentiate one encryption algorithm from another its ability to secure the protected data against attacks by hacker and its speed and efficiency. This paper provides a performance comparison between five of the most common encryption algorithm. This paper also presents a highly efficient implementation of NTRU.

From the survey of the various existing systems it is concluded that whether there are so many methods or algorithms which are used for the security purpose in the distributed database system for file sharing but still they have some problems, to overcome these problems the NTRU is considered as the best algorithm for the security purpose.

Sr. no	**Author name**	**Work done**
1	Gurkamal Bhullar	analyze the NTRU algorithm, based Encryption decryption technique for security to Distributed Database System.
2	Luca Ferretti	describes the architecture that integrates cloud database services with data confidentiality and the possibility of executing the concurrent operations on encrypted data.

3	L. Ferretti	Defines architecture that combines data encryption, key management, authentication and authorization solutions, and that addresses the issues related to typical threat scenarios for cloud database services.
4	Ismail Omar Hababeh	presents an intelligent clustering method for distributed database system that provides a structure for organizing large number of network sites into a set of useful clusters to minimize transactions processing communications
5	Yashpal Mote	provides a performance comparison between five of the most common encryption algorithm. This paper also presents a highly efficient implementation of NTRU.

Table no. 1 Comparison of the existing systems by various authors.

3. Cryptographic Techniques

Cryptography is the study of Secret (crypto-) and Writing (- graphy). It is the science or art of encompassing the principles and methods of transforming an intelligible message into one that is intelligible and then transforming the message back to its original form. Today’s cryptography is more than encryption and decryption. Authentication is as fundamentally a part of our lives as privacy. We use authentication throughout our everyday lives when we sign our name to some document and for instance, as we move to world where our decisions and agreements are communicated electronically, we need to have electronic techniques for providing authentication.

There are three types of cryptographic techniques:

i. Symmetric-key cryptography.
ii. Asymmetric-key cryptography.
iii. Hash function cryptography.

i Symmetric-key cryptography- refers to encryption methods in which both the sender and receiver share the same key. Because of the same key the encryption-decryption process is fast but more security risks are here. Symmetric-key encryption can use either stream ciphers or block ciphers. Stream ciphers encrypt the digits (typically bytes) of a message one at a time. Block ciphers take a number of bits and encrypt them as a single unit, padding the plaintext so that it is a multiple of the block size. Blocks of 64 bits have been commonly used. The Data Encryption Standard (DES) and the Advanced Encryption Standard (AES) are block cipher designs which have been designated cryptography standards by the US government (though DES's designation was finally withdrawn after the AES was adopted). Despite its deprecation as an official standard, DES (especially its still-approved and much more secure triple-DES variant) remains quite popular; it is used across a wide range of applications, from ATM encryption to e-mail privacy and secure remote access. Many other block ciphers have been designed and released, with considerable variation in quality and some have been thoroughly broken.

ii. Asymmetric-key cryptography:- Public-key cryptography, also known as asymmetric cryptography, is a class of cryptographic protocols based on algorithms that require two separate keys, one of which is secret (or private) and one of which is public. It is computationally easy for a user to generate a public and private key-pair and to use it for encryption and decryption. Although different, the two parts of this key pair are mathematically linked. The public-key cryptography is more secure because different keys are used for encryption and decryption. The various asymmetric key algorithms are RSA, Diffie–Hellman key exchange and Digital Signature Algorithm.

iii. Hash Functions-Hash functions, also called message digests and one-way encryption algorithms that, in some sense, use no key .Instead, a fixed-length hash value is computed based upon the plaintext that makes it impossible for either the contents or length of the plaintext to be recovered. Hash algorithms are typically used to provide a digital fingerprint of a file's contents often used to ensure that the file has not been altered by an intruder or virus. Hash functions are also commonly employed by many operating systems to encrypt passwords. Hash functions, then, provide a measure of the integrity of a file. Mostly MD5 and SHA hash function cryptography are used for security purpose. Table no.2 shows the comparison of some cryptographic techniques.

Some commonly used encryption algorithms are:

DES: (Data Encryption Standard), was the first encryption standard to be recommended by NIST (National Institute of Standards and Technology). It is based on the IBM proposed algorithm called Lucifer. DES became a standard in 1974. Since that time, many attacks and methods recorded that exploit weaknesses of DES, which made it an insecure block cipher.

3DES: As an enhancement of DES, the3DES (Triple DES) encryption standard was proposed. In this standard the encryption method is similar to the one in original DES but applied 3 times to increase the encryption level. But it is a known fact that 3DES is slower than other block cipher methods.

AES: (Advanced Encryption Standard), is the new encryption standard recommended by NIST to replace DES. It was originally called Rijndael (pronounced Rain Doll). It has variable key length of 128, 192, or 256 bits; default 256. AES encryption is fast and flexible; it can be implemented on various platforms especially in small devices.

Table no. 2 comparison of cryptographic techniques

Parameters	**Symmetric key**	**Asymmetric key**	**Hash functions**
Speed up	Fastest	Slow	Fast
Collision resistant	No	No	Yes
Complexity	Less	More complex	Less than asymmetric
Key agreement/ exchange	A big problem	No problem at all but this technique requires a lot of mathematical calculations	No such problem at all
Delays	Less	More	Less than asymmetric
Security	Easily breakable	Better than symmetric key	Impossible to break
Software implementations	Difficult Requires tables and complex programs	Difficult Requires tables and complex programs	Simple donot requires tables and complex programs

NTRU: The NTRU Encryption schemes is an interesting alternative to well established schemes such as ,RSA,DES, ElGamal, and ECIES. The security of NTRU depends on the hardness of computing short lattice vectors and thus is a promising user for being quantum computer resistant. There has been extensive research on efficient implementation of the NTRU encryption scheme.

4. ENCRYPTION ANALYSIS OF DES, RSA, NTRU

In distributed database we work upon the concurrency and security, and our main objective behind this work is to provide more security to our data with the help of the NTRU algorithm [12], because it is the latest algorithm being introduced in the year 2009 and is approved by the institute of electrical engineers. This work is mainly done by the three mathematicians Jeffrey Hoff stein, Joseph H Silverman, Jill Pipher in the year 1996 and later at the end of 1996 these mathematicians plus Daneil founded the NTRU cryptosystem, and on 2009 it was approved. In the later part of this we are analyzing that performance comparison between mostly used algorithms that are DES and RSA with NTRU.

Points for discussion	**Identification Based Model**	**File encryption based Model**	**Secured channel using model**	**Distributed Server Based architecture**
Ways of ensuring security	Only identify the authorized person, so hacker can get access on database	Key & file both remains in one server. So, getting access on one server helps to get all information	Intruder cant access the data, but uploaded file is not secured	Ensures security in data exchanging process. Only getting control over full system can leak information
Information leakage probability	Medium	Medium	Medium	Low
Execution time	Small	Medium	Small	Medium
Security Breaking probability	Medium	Medium	Medium	Probably Low than others

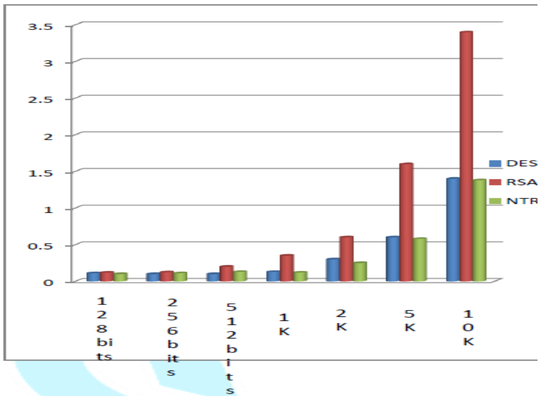

Fig: 1.4 Encryption Analysis

The above graphical representation clearly identifies that as the no of bytes goes on increasing the time to encrypt the data also increases in all the three cases but the NTRU needs less time in encryption.

Table no.3 Comparison of existing models

The table no.3 represents the comparison of the various existing models which are used for the data security purpose in the distributed database system environment and the cloud computing environment.

5. CONCLUSION

In our work we mainly deal with the distributed database, and in that we try to solve the concurrency and security. Concurrency means when more than one transaction were executed at the same by the user, like if more clicks on a particular site occur then traffic increases and this consequences will result the server crash, to avoid this problem proper measures were taken in order to avoid it. Security plays important role in this work as to protect our sensitive information from the unauthorized user this will be conducted with the help of certain algorithm and in that algorithm we will try to adopt the asymmetric approach. By NTRU being a strong encryption technique will surely provide the strong security as till date its has no security clauses which will prove to be a best fitted approach for our research thesis.

6. REFERENCES

[1]. Gupta V.K., Sheetlani Jitendra, Gupta Dhirajand Shukla Brahma, (2012), "Data concurrency control and security issues of distributed database transaction" *NIMS University, Jaipur, Rajasthan, INDIA*, 1(2), pp:70-73.

[2]. Amandeep Kaur Gill, Charanjit Singh, (2014), " Security of N-Tier Architecture using

NTRU" , *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(7), pp:1018-1022.

[3]. Dr. Lokanatha C. Reddy,(2011), "A Review on Data mining from Past to the Future", *International Journal of Computer Applications*,15(7).

[4]. Gurkamal Bhullar, Navneet Kaur, (2014), "Concurrency and Security Control with NTRU", *International Journal of Innovative Research in Computer and Communication Engineering*, 2(3), pp:3352-3357.

[5] Ismail Omar Hababeh (2010), "Intelligent Network Communications for Distributed Database Systems" , *Second International Conference on Advances in Databases, Knowledge, and Data Applications*, pp:69-74.

[6]. Luca Ferretti, Michele Colajanni, and Mirco Marchetti (2014), "Distributed, Concurrent, and Independent Access to Encrypted Cloud Databases", *Institute of Electrical and Electronics Engineers (IEEE)*, 25(2), pp:437-446.

[7]. L. Ferretti, M. Colajanni, (2012), "Supporting Security and Consistency for Cloud Database," *Proc. Fourth Int'l Symp. Cyberspace Safety and Security.*

[8] Luca Ferretti, Fabio Pierazzi, Michele Colajanni, and Mirco Marchetti, (2014), "Scalable Architecture for Multi-User Encrypted SQL Operations on Cloud Database Services", *Institute of Electrical and Electronics Engineers (IEEE)*, 2(4), pp:485-498.

[9]. Ran Vijay Singh and M.P.S Bhatia, (2011), "Data Clustering with Modified K-means Algorithm", *IEEE International Conference on Recent Trends in Information Technology*, pp 717-721.

[10]. Sheetlani Jitendra, Gupta V.K, (2012) "Concurrency Issues of Distributed Advance Transaction Process*", Res. J. Recent Sci.* , 1(2), pp:426-429.

[11]. S.Mekala, M.Senthil Prabhu M.E, (2014) , "Survey on Encrypted Database in Cloud", *International Journal of Advance Research in Computer and Communication Engineering*, 3(10), pp:8290-8293.

[12]. Subedari Mithila, P. Pradeep Kumar, (2011), "Data Security through Confidentiality in Cloud Computing Environment", *(IJCSIT) International Journal of Computer Science and Information Technologies*, Vol. 2, pp:1836-1840.

[13].Yashpal mote and Shekhar Gaikwad, (2012), "Superior security data encryption algorithm*" International journal of engineering science*, vol-6.