Privacy Preservation and Data Leak Prevention in Cloud Database

Raunak Dubey, AmitSaxena

Truba Institute of Engineering and Information Technology, Bhopal, M.P. India

Abstract— Cloud computing is a set of rules for permitting an appropriate, universal, on-demand network access. Cloud processing is the outcome of development and acceptance of current technologies and prototypes. Massive progression in digital information and data, improved broadband conveniences, varying data storage necessities, and Cloud system computing supervised to the appearance of cloud databases. The information and data should be preserved and protected. Detecting and preventing data privacy involves a fixed of different technique, which may include data-privacy recognition, surreptitious malware detection data locking up, and policy enforcement. The dissimilar safety problems in cloud are Concealment, scalability, data truthfulness, heterogeneity, data intrusion, Non- Disclaimer, access control, authentication and authorization. Confidentiality of information data is additional safety issue connected with cloud computing environment. This research paper propose a unique cloud system for privacy preservation and data leak prevention of sensitive data. In this system we recommend an architecture which provides cloud database security for useful information

Keywords— Cloud Database,Security, Privacy Preservation, Audit, Data leak Detection

I. INTRODUCTION

Cloud computing[1] is a fixed of rules for permitting an appropriate, universal, on-demand network access. At its simplest form, cloud processing is the dynamic delivery of information knowledge capabilities and resources as a facility above the Internet. The cloud processing resources are storage, networks, applications, servers, and services. This cloud computing technology is involves of five important features, three service archtypes, and four deployment simulations [2]. Enormous succession in digital information and data, better broadband conveniences, altering data storage provisions, and Cloud computing led to the appearance of cloud databases. An essential objective[3] of cloud processing technology is to make available on-demand admittance to computational resources such as network, database, applications and platform on pay-as you-go basis alike to the way in which we get services from public usefulness services such as electricity, water, telephony, and gas. Cloud clients can be either application /software service providers. A cloud source is a vendor or company that provides economically effective cloud services using the h/w(hardware) and s/w(software).

There are four main categories of cloud processing facility models. These models are DaaS-Dbms-as-a-Service, SaaS-Software as a Service, IaaS-Infrastructure as a Service and PaaS-Platform as a Service. Dbms-As-A Service (DAAS)[3]-Cloud database is aimed for virtualized computer environment. It is none as modest as taking relational dbmsmgmt and positioning it over a cloud database server. Platform as a service (PaaS)-In platform as a facility ideal, service provider offers computer devices called hardware and software programs called as software to the consumer which is looked-for by him to database and web server. S/w-as-a-service (SaaS)-S/waas can be well-defined as the computer programs called software that is positioned over the internet. Infrastructure as a service (IaaS) is the furthermost rudimentary cloud facility prototypical. It delivers computers virtual machines physical or and other possessions.

Regardless of the above revealed service models, cloud services can be set up in four ways liable upon the consumers' necessities. Community Cloud, Private Cloud, Hybrid Cloud and Public Cloud. Cloud computing arrangement, set accessible only to a particular client and managed either by the association itself or third party facility source is called private cloud. A cloud association is provided to many clients and is accomplished by a third party is called a public cloud. Infrastructure shared by numerous establishments for a shared cause for particular community is called community cloud computing. An arrangement of two or extra additional cloud distribution models is called hybrid cloud. Virtualization and cloud computing technology are used interchangeably in many situation, but this is inappropriate. Virtualization has develop a practical requirement these days, and the tendency is ongoing for a decent motive because when applied, it delivers many profits such decrease in operational costs and capital, Hard-to-find person's resource savings, Physical space decrease, Access to storage resources, network, server and on demand, Energy savings for a greener environment. Cloud processing structural design consists of end users, Internet and cloud providers. The end users can be mobile devices, applications and different computer software's, Internet with high speed, and different cloud service providers. PaaS providescomputers physical or virtual machines and

further resources. The providers like Amazon[4], Rackspace are the examples of infrastructure as a service. The information and data should be preserved and protected. The information should not be visible to anyone at any cost. Confidentiality of information data is additional safety issue connected with cloud computing upbringing. The different safety measures in cloud are scalability, Non- Repudiation, Confidentiality, access control, heterogeneity, Data Intrusion[5], Data Integrity, authentication and authorization. Although data encryption[5] appears the ultimate in-built way out for data privacy. For performance end of vision the SQL statements must be executed without decryption. Some solution download the entire cloud database into native place and decrypt it. After decryption it execute SQL statements and encrypt it to store up data in cloud database. But this process have some performance problems. Key circulation and key storing are more challenging issue in the cloud database. Cloud information stowage is a virtual database storage that certificates customers to store up entities and documents. Cloud database should provision of cloud processing with traditional relational databases for extensive satisfactoriness. Improving the secrecy of statistics and figures stowed in cloud-computing databases signifies an important involvement to the acceptance of the cloud-computing as the fifth usefulness because it addresses most user apprehensions. The probable encounters connected with cloud database are high availability, scalability and error acceptance. The further encounters are data consistency and truthfulness, secrecy and many more.

The paper is organized as follows. Section 2 represent the literature survey. Section 3 represent the projected work and various steps in proposed work. Division 4 provides the implementation detail of the proposed system. Section 5 concludes the paper with future planning.

II LITERATURE SURVEY

Attribute-based encryption (ABE) [6] Attributes are the properties of user which represent various values related to profile. It permits every cipher text to be connected with an attribute. The system consists of master undisclosed key for attributes. Depending on the policy of attributes master key holder get the secret key and decrypt the desired data from database. The main anxiety in ABE is complicity resistance but not the compression of secret keys. The main issue is range of the undisclosed key in ABE. Definitely, the volume of the key frequently rises linearly with the amount of attributes.

Proxy re-encryption[7] is to improve the decryption power. PRE is mainly used to delegate the decryption keys of cipher texts without circulation of the top-secret key to the receiver. A PRE system allows sender to over hand to the cloud database server the competence to translate the cipher texts encrypted to receiver. PRE is fine and known to have many applications including cryptographic file system.

Predefined Hierarchy based Cryptographic Keys:

Cryptographic key consignment systems [8] intention to minimalize the overhead in managing and storing secret keys for universal cryptographic use. In this method a secret key for a given node can be used for its sibling's nodes. The author in this projected effort delivers a method using hierarchical system. This scheme uses tree structure which consists of nodes. The top-secret key is assigned to the parent node the all nodes under parent node automatically grants the top-secret keys to all other nodes. This scheme produce a tree grading of symmetric-keys with the help of recurring assessments of pseudorandom block cipher on an immovable secret. The concept be capable to be used in graph. Additional innovative cryptographic key consignment systems provision access strategy that can be demonstrated by cyclic graph and an acyclic graph. Maximum of these systems generate secrets aimed at symmetric-key cryptosystems. This structure is more expensive then symmetric keys.

Key circulation and key storing are more challenging issue in the cloud database. Cloud storage is a virtual storage that enables users to store up documents and objects. Cloud database should provision of cloud computing and traditional relational databases for extensive satisfactoriness. The probable encounters connected with cloud database are high availability, scalability and error acceptance. The further encounters are data consistency and truthfulness, secrecy and a lot more. [9] propose a secure architecture using encryption and single user key distribution for cloud database. Improving the secrecy of statistics and figures stowed in cloud processing databases signifies an important involvement to the acceptance of the cloud processing as the fifth usefulness because it addresses most user apprehensions.

Identity-based Encryption (IBE) by Compact Keys:

Identity-based Encryption is a category of public key encryption. [10] tried to construct IBE with key accumulation. IBE scheme also uses random oracle.

In this IBE scheme, key combination is inhibited in the all keys should be taken from dissimilar "identity divisions." The benefit of this schemes are security and cipher text size.

Although data encryption appears the ultimate in-built way out for data privacy. For performance fact of vision the SQL statements must be executed without decryption. Some solution download the entire cloud database into

native place and decrypt it. After decryption it execute SQL statements and encrypt it to store up data in cloud database. But this process have some performance problems.

The correct auxiliary footstep is to perform SQL methods on the cloud processing database, without providing decryption keys to the cloud database provider. An early resolution obtainable in [11] is founded on data combination systems [12]. A scheme associate plain text meta data to group of encrypted data. Though, plain text meta data may disclose subtle statistics and figures combination presents pointless network overheads.

III PROPOSED METHOD

The proposed work consists of following steps.

Step 1: Data owner Application

Step 2: Security key generation

Step 3: Security key distribution

Step 4: Data auditing process

Step 5: Data auditing report generation

Step 6: Data leak detection

Step 7: Data leak alert

Step 8: Data leak report

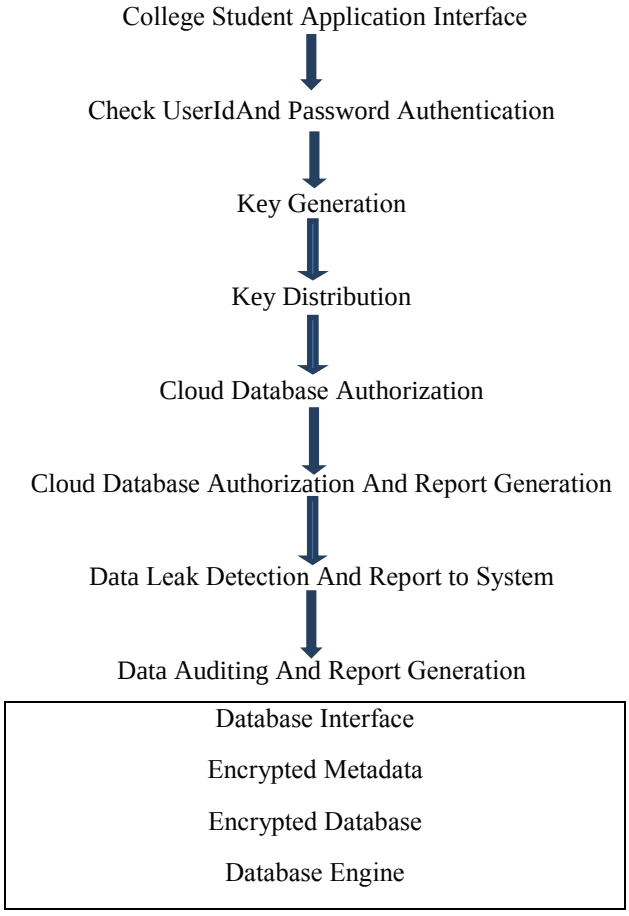

Figure Flowchart of proposed system architecture

The proposed architecture consists of clients, intermediate servers, and cloud database. The client can be mobile client, desktop computer etc. The intermediate servers are included into architecture to provide higher level of security. The database servers are used to store database of organizations. The user be able to interact with the cloud database using application. The flow chart shown below represent proposed steps.

Figure: Projected System Structural design

Master key generation: In immediate step the system generate the master key which is used for authentication purpose. The immediate step is to generate multiuser key which is used for various groups for security purpose. In immediate step is distribution of the multiuser key to other user participating in cloud database services. After getting the multiuser security key the user can access cloud database and can execute different SQL statements and get the desired information. The authentication is performed using host Internet Protocol Address and userid/ password provided by administrator. After authentication user authorization is also performed using intermediate server. The audit is performed by intermediate server and report is generated and provided to administrator. The data leak recoginition of susceptible information is performed by intermediate server and report is given to the administrator to improve security. The proposed model can improve the performance of the cloud database. All the data stored in the cloud database are in encrypted form. The application designed for testing the proposed architecture can execute SQL statements select, insert, delete, and update to the encrypted database.

The proposed work also implement traceability i.e. ability to group manager or original user to preserve the identity.

It also provides privacy preservation to the cloud database. The data leak detection is also provided in this paper.

IV. IMPLEMENTATION

Java platform is used for the execution of the algorithm and Oracle 11g server is choosen as the back-end. Windows operating-system is considered good in the security point of view. The implementations are carried out in lab, which make available with a cluster of machines in Oracle 11g database and Java environment and programming language. Every client computer executes the Java environment client prototype of structural design on a Intel PIV machine having a single 3 GHz processor, 2 GB of RAM and two 7200 RPM 500 GB SCSI disks. The database server is Oracle 11g running on Intel machine having a PIV 3.5 GHz processor, 4GB of RAM and a 7,200 RPM 500 GB SATA disk. The implementation is tested with 4, 10, 15 and 20 client machines. The database used for experiment is college training and placement database. We have collected training and placement data from college of different years. We have also collected various company data in which students are placed. The database column have number, varchar2 and date data type. The implemented system supports all basic SQL operations like insert, alter, select, delete with where clause. Our system also supports integrity constraints, some SQL basic functions and procedures.

V. RESULT ANALYSIS

The figure below shows the throughput of the system with 5, 10, 15 and 20 clients. The throughput is evaluated with plaintext database and encrypted database. As represents in figure the throughput of plaintext result is very much closed to throughput of encrypted database result. As in figure 1, figure 2, figure 4 and figure 4 transactions per minute is very closed to for latencies higher than 80ms for all possible combinations of 5, 10, 15 and 20 clients and network latency of 0 to 120ms. This results demonstrate that the system is useful for cloud database.

The overheads of the performance and data confidentiality for cloud database services are discussed. The performance tests will carried out to estimate the throughput for increasing number of clients and different network latencies.

VI. CONCLUSIONS

The cloud computing resources are storage, networks, applications, servers, and services. Database As A Service (DAAS)- Cloud database is aimed for virtualized processor system. The cloud system database as a service is a pioneering prototype that can support numerous Internet-based applications. The Cloud providers also called service producers and Cloud customers also called service consumers or clients are the main pillars in cloud database. The prospective encounters connected with cloud system database are high availability, scalability data consistency and fault tolerance, integrity, reliability and many more. Although data encryption appears the ultimate in-built way out for data privacy.The information and data should be preserved and protected. The cloud database should be secure in 3-terms of audit, authentication and authorization. Confidentiality of information data is additional safety issue connected with cloud computing environment. We have design a host assisted privacy preservation cloud database which ensures security of the information from not permitted access. The proposed architecture keep large organization cloud database from data leak and misuse. The experimental result illustrated that our system is fine appropriate for secrecy safeguarding and data leak prevention in cloud database. We have developed a secure cloud database architecture for privacy preservation of large organization and susceptible data leak detection and report to the system.

A direction of future investigation we are planning to introduce iris detection authentication system to advance the security in our work. In are also planning to implement our system in public cloud.

REFERENCES

[1] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic,“Cloud computing and emerging it platforms: Vision, hype, andreality for delivering computing as the 5th utility,” Future GenerationComput. Syst., vol. 25, no. 6, pp. 599–616, 2009.

[2] H. Hacig€um€u¸s, B. Iyer, and S. Mehrotra, “Providing database as a service,” in Proc. 18th IEEE Int. Conf. Data Eng., Feb. 2002, pp. 29–38.

[3] Peter Mell,TimothyGrance,”The NIST definition of cloud computing”,http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

[4] D. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in Proc. IEEE Symp. Security Privacy, May 2000, pp. 44–55.

[5] R. A. Popa, C. M. S. Redfield, N. Zeldovich, and H. Balakrishnan, “CryptDB: Protecting confidentiality with encrypted query processing,” in Proc. 23rd ACM Symp. Operating Systems Principles, Oct. 2011, pp. 85–100.

[6] C. Rathgeb and A. Uhl, “Privacy Preserving Key Generation for Iris Biometrics,” Proc. 11th IFIP TC 6/TC 11 Int’l Conf. Comm. And Multimedia Security, pp. 191-200, 2010.

[7] Privacy-Preserving Detection of Sensitive Data Exposure, XiaokuiShu, Danfeng Yao, and Elisa Bertino, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 10, NO. 5, MAY 2015, pp-1092-1112

[8] Yan Sui, XukaiZou, Eliza Y. Du, Feng Li,"Design and Analysis of a Highly User-Friendly,Secure, Privacy-Preserving, and RevocableAuthentication Method ",IEEE TRANSACTIONS ON COMPUTERS, VOL. 63, NO. 4, APRIL 2014, pp-902-916

[9] J. Daugman, "How Iris Recognition Works," IEEETrans. Circuitsand Systems for Video Technology, vol. 14,no. 1, pp. 21-30,Jan. 2004.

[10] J. Dai, J. Feng, and J. Zhou, "Robust and Efficient Ridge-BasedPalmprint Matching," IEEE Trans. Pattern Analysis and MachineIntelligence, vol. 34, no. 8, pp. 1618-1632, Aug. 2012.

[11] Z. Zhou, E. Du, N. Thomas, and E. Delp, "A New Human IdentificationMethod: Sclera Recognition," IEEE Trans. Systems, Man, andCybernetics, Part A: Systems and Humans, vol. 42, no. 3, pp. 571-583,May 2012.

[12] A. Ahmed and I. Traore, "A New Biometric Technology Based onMouse Dynamics," IEEE Trans. Dependable and Secure Computing,vol. 4, no. 3, pp. 165-179, July/Sept. 2007.