# An Efficient Public Auditing with Privacy Preserving Identity and Traceability

Nitesh Kumar Namdeo
Department of Computer Science and Engineering,
RGPV, India

Sachin D. Choudhari, PhD
Department of Computer Science and Engineering,
RGPV, India

## ABSTRACT

Cloud computing is a model for enabling ubiquitous, 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. Cloud computing is a general term for the delivery of hosted services over the internet. Cloud computing enables companies to consume the resources and compute their utility rather than building and maintaining computing infrastructure. A cloud database is a database that has been optimized or built for a virtualized computing environment. Since these data-centers may be located in any part of the world beyond the reach and control of users, there are multifarious security and privacy challenges that need to be understood and addressed. Cloud has been prone to various security issues like storage, computation and attacks like Denial of service, Distributed Denial of Service, Eavesdropping, insecure authentication or logging etc. Privacy preservation is main security issue in public cloud. This paper proposed architecture for privacy preservation and traceability. The implementation results represent that our method is suitable for large organizations.

## Keywords

Cloud database, Security, Privacy Preservation, Auditing, Authentication, DaaS

## 1. INTRODUCTION

### A. Overview

Cloud computing can be defined as new computational capabilities that focus on both academia and industry. Cloud computing is the outcome of development and acceptance of current technologies and prototypes. Cloud computing is a prototype for permitting universal, appropriate, on-demand network access. The cloud computing resources are storage, networks, applications, servers, and services. Cloud computing is a style of computing where enormously scalable IT-enabled proficiencies are delivered ‘as a service’ using Internet technologies to multiple outdoor clients. At its simplest form, cloud computing is the dynamic delivery of information technology resources and capabilities as a service over the Internet. This cloud computing model is consists of five important features, three service prototypes, and four deployment models [1]. Enormous progression in digital information and data, better broadband conveniences, altering data storage necessities, and Cloud computing led to the appearance of cloud databases.

Massive growth in digital data[2], changing data storage requirements, better broadband facilities and Cloud computing led to the emergence of cloud databases .Cloud Storage, Data as a service (DaaS)[3] and Database as a service (DBaaS) are the different terms used for data management in the Cloud. They differ on the basis of how data is stored and managed. Cloud storage is virtual storage that enables users to store documents and objects. Drop box, iCloud[3] etc. are popular cloud storage services. DaaS allows user to store data at a remote disk available through Internet. Cloud storage cannot work without basic data management services. So, these two terms are used interchangeably. DBaaS is one step ahead. It offers complete database functionality and allows users to access and store their database at remote disks anytime from any place through Internet. Amazon’s Simple DB, Amazon RDS, Google’s BigTable, Yahoo’s Sherpa and Microsoft’s SQL Azure Database are the commonly used databases in the Cloud. The data should be kept secured and should not be exposed to anyone at any cost. Confidentiality of data is another security issue associated with cloud computing. The different security issues in cloud are scalability, heterogeneity, Data Intrusion[4], Data Integrity, Non- Repudiation, Confidentiality, access control, authentication and authorization. Based on many discussions with customers and surveys, the following security and integration issues seem to be on many customers’ minds: How the cloud will keep data secure and available? How to comply with current and future security and risk management compliance? What type of security services are available through the cloud? How to perform internal and external audits of cloud security? How to automate network, compute, and storage provisioning? 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 environment. Although data encryption[5] appears the ultimate in-built way out for data privacy. For performance point of view 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 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 storage is a virtual storage that enables users to store 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 fault tolerance. The other challenges are data reliability and truthfulness, confidentiality and many more.

### B. Motivation

Cloud services, similar to many other services, are perishable in nature and cannot be stored for future sale. The cloud database as a service is an innovative prototype that can support numerous Internet-based applications. The main problem for adoption for cloud database is the information privacy complications. The objective of cloud computing is to permit users to take advantage from all of cloud related technologies, without the essential knowledge about or proficiency with each one of them. The different security issues in cloud are scalability, heterogeneity, Data Intrusion, Data Integrity, Non- Repudiation, Confidentiality, access control, authentication and authorization. The different encryption techniques that allow the execution of SQL operations on encoded data have some performance limits. And different types of encryption techniques must be implemented for every database SQL operation and database column. Most of the encryption technique regarding encryption for cloud-based database services are inappropriate to the database prototype. The number of leaked confidential data records has increased throughout the past few years. Detecting and preventing data privacy requires a set of different technique, which may include data-privacy recognition, surreptitious malware detection data locking up, and policy enforcement. Intentionally scheduled attacks, unintended leaks such as accelerating trustworthy emails to unclassified email account. The attacks may also include human faults such as passing on the erroneous privilege main reason of the data-privacy occurrences. Database as a service (DBaaS) that poses several research challenges in terms of security and cost evaluation from a tenant’s point of view.

Paper is organized as follows. Section II provides literature survey of privacy preservation and security. Section III provides the proposed architecture and algorithm. Section IV represents the implementation of proposed work. Section V concludes the paper.

## 2. LITERATURE SURVEY

Several mechanisms have been designed to allow both data owners and public verifiers to efficiently audit cloud data integrity without retrieving the entire data from the cloud server. However, public auditing on the integrity of shared data with these existing mechanisms will inevitably reveal confidential information—identity privacy—to public verifiers. Cloud service providers offer users efficient and scalable data storage services with a much lower marginal cost than traditional approaches [2]. It is routine for users to leverage cloud storage services to share data with others in a group, as data sharing becomes a standard feature in most cloud storage offerings, including Dropbox, iCloud and Google Drive.

Recently, many mechanisms [5] ] have been proposed to allow not only a data owner itself but also a public verifier to efficiently perform integrity checking without downloading the entire data from the cloud, which is referred to as public auditing [6]. In these mechanisms, data is divided in to many small blocks, where each block is independently signed by the owner; and a random combination of all the blocks instead of the whole data is retrieved during integrity checking [7]. A public verifier could be a data user (e.g., researcher) who would like to utilize the owner’s data via the cloud or a third-party auditor (TPA) who can provide expert integrity checking services [12]. Moving a step forward, [13] designed an advanced auditing mechanism [8], so that during public auditing on cloud data, the content of private data belonging to a personal user is not disclosed to any public verifiers. Unfortunately, current public auditing solutions mentioned above only focus on personal data in the cloud.

Provable data possession (PDP), proposed [13], allows a verifier to check the correctness of a client’s data stored at an un-trusted server. By utilizing RSA-based homomorphic authenticators and sampling strategies, the verifier is able to publicly audit the integrity of data without retrieving the entire data, which is referred to as public auditing. Unfortunately, their mechanism is only suitable for auditing the integrity of personal data. [9] defined another similar model called Proofs of Retrievability (POR), which is also able to check the correctness of data on an un-trusted server. The original file is added with a set of randomly-valued check blocks called sentinels.

The verifier challenges the un-trusted server by specifying the positions of a collection of sentinels and asking the un-trusted server to return the associated sentinel values.

[10] designed two improved schemes. The first scheme is built from BLS signatures, and the second one is based on pseudo-random functions. To support dynamic data, [15] presented an efficient PDP mechanism based on symmetric keys. This mechanism can support update and delete operations on data, however, insert operations are not available in this mechanism. Because it exploits symmetric keys to verify the integrity of data, it is not public verifiable and only provides a user with a limited number of verification requests.

[13] utilized Merkle Hash Tree and BLS signatures to support dynamic data in a public auditing mechanism. [14]introduced dynamic provable data possession (DPDP) by using authenticated dictionaries, which are based on rank information. [15]exploited the fragment structure to reduce the storage of signatures in their public auditing mechanism. In addition, they also used index hash tables to provide dynamic operations on data. The public mechanism proposed by [12] are able to preserve users’ confidential data from a public verifier by using random markings. In addition, to operate multiple auditing tasks from different users efficiently, they extended their mechanism to enable batch auditing by leveraging aggregate signatures [13].

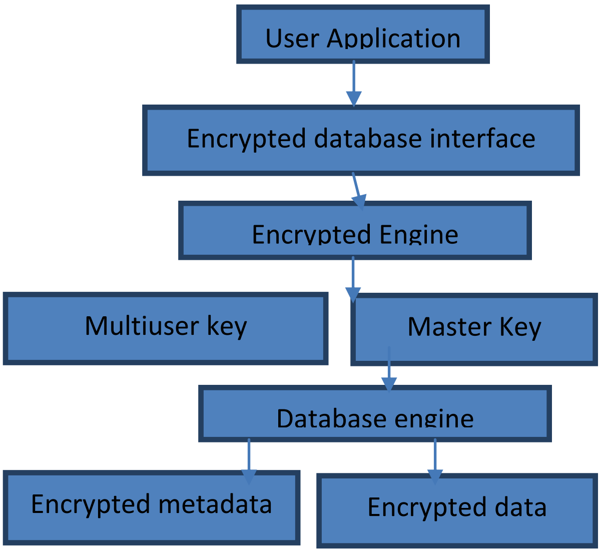
[13] leveraged homomorphic tokens to ensure the correctness of erasure codes-based data distributed on multiple servers. This mechanism is able not only to support dynamic data, but also to identify misbehaved servers. To minimize communication overhead in the phase of data repair,[14] also introduced a mechanism for auditing the correctness of data under the multi-server scenario, where these data are encoded by network coding instead of using erasure codes. More recently, [15] constructed an LT codes-based secure and reliable cloud storage mechanism. Compare to previous work [13], [14], this mechanism can avoid high decoding computation cost for data users and save computation resource for online data owners during data repair.

Oruta[6], a privacy-preserving public auditing mechanism for shared data in the cloud. We utilize ring signatures to construct homomorphic authenticators, so that a public verifier is able to audit shared data integrity without retrieving the entire data, yet it cannot distinguish who is the signer on each block. To improve the efficiency of verifying multiple auditing tasks, [6] further extend our mechanism to support batch auditing. Propose a novel privacy-preserving mechanism that supports public auditing on shared data stored in the cloud. In particular, exploit ring signatures to compute verification metadata needed to audit the correctness of shared data. With our mechanism, the identity of the signer on each block in shared data is kept private from public verifiers, who are able to efficiently verify shared data integrity without retrieving the entire file. In addition, our mechanism is able to perform multiple auditing tasks simultaneously instead of verifying them one by one.

# 3. PROPOSED WORK

In previous data auditing methods entire database is downloaded from the cloud database providers. Then after performing the auditing process upload the entire database to the cloud database. In general the size of cloud database is huge. So downloading and uploading the entire database is time consuming and also utilizes computing resources such as CPU time, memory and network. The cost is also improved with this technique. During data verification process the confidential data or personal data will not be disclosed to any integrity checking authorities. Intermediate server provides scalability and availability of cloud database servers.

The multi-user key distribution architecture is not supported by the previous work. Here multi-user key distribution scheme is proposed for the cloud database which will use the encryption algorithm. Encryption scheme in database has performance limits and different technique for different SQL operation. The proposed encryption technique encrypts each plain column to numerous encrypted columns. The encrypted value is encapsulated in dissimilar layers of encryption. In this scheme external layers assurance higher privacy but support less computations competences with respect to inner layers. Our architecture consist of 3 layers. Clients, Intermediate servers, Database servers. Intermediate servers provide availability and scalability of cloud service. The algorithm that will be designed for implementation will consist the following steps. Master key generation, Multi-user key generation, Multi-user key distribution, Database creation, Execution of SQL operations and Reducing cost using cost pricing model.



Master key generation: In first step the system generate the master key which is used for authentication purpose.

The next step is to generate multiuser key which is used for various groups for security purpose.

In next 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 proposed model can improve the performance of the cloud database. The proposed architecture can also reduce the cost of the services with the help of proposed cost model.

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. Data and information transferred between cloud database and application are not encrypted. The data and information always encrypted before storing into the cloud database. When an application issue SQL commands insert, select, update, and delete the encrypted database interacts with encryption engines with the help of master key. After getting the multiuser security key the user can access cloud database and can execute different SQL statements and get the desired information.

## 3.1 Database creation

The database administrator setup the database. In database creation phase database administrator creates the database with required tables. The column of a tables may have number, varchar and date data types. The database administrator insert new rows into the table. The database administrator distributes the master key to all the cloud database clients.

SQL statements execution:

SQL statements insert, delete, select and update are executed on database. When an application or user executes SQL operations on the cloud database encryption engine determines the table, SQL operations and column of database and decrypt the desired result and provide the output data to the clients. The application or user uses master key for decryption process.

# 4. IMPLEMENTATION AND RESULTS

Java platform is used for the implementation of the algorithm and Oracle 11g server is used 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, update, select, delete with where clause. Our system also supports integrity constraints, some SQL basic functions and procedures. We have tested our system with plaintext data as well as encrypted data. Plaintext data means database having values without encryption. In encrypted database all the columns are stored as encrypted form for security purpose. We have performed the test for every database i.e. for plaintext and encrypted. The test is performed for plaintext and encrypted database for different clients from 5 to 20. The network latencies are increased from 0 to 120ms. The experiments are tested to evaluate the overhead of database by plaintext and encrypted data. The experiments also test the response time and throughput. In experiments the transactions are performed and response time and throughput are recorded with different number of clients and different time interval.

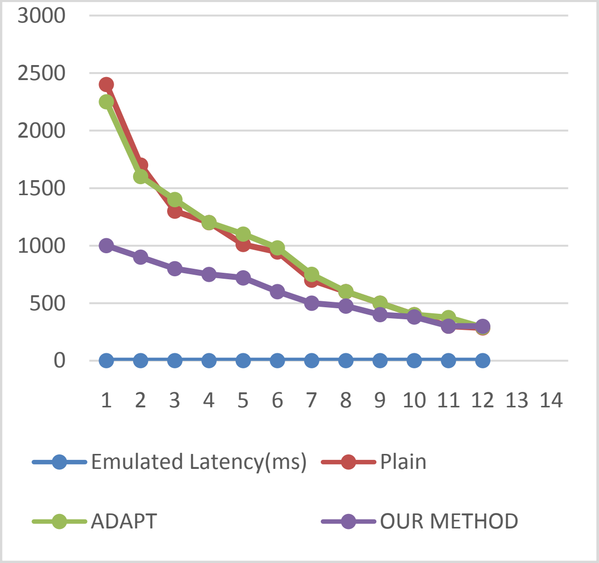


**Fig. 1 Throughput with 20 clients**

The figure below shows the throughput of the system with 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. This results demonstrate that the system is useful for public cloud database.

## 5. CONCLUSION

Cloud computing is a computing in which large groups of remote servers are networked to allow centralized data storage and online access to computer services or resources. Several mechanisms have been designed to allow both data owners and public verifiers to efficiently audit cloud data integrity without retrieving the entire data from the cloud server. However, public auditing on the integrity of shared data with these existing mechanisms will inevitably reveal confidential information—identity privacy—to public verifiers. The major issue with the cloud database is that it requires a very high level security. Data are not always safe when they are stored inside cloud providers. Since these data-centers may be located in any part of the world beyond the reach and control of users, there are multifarious security and privacy challenges that need to be understood and addressed. Privacy preservation is main security issue in public cloud. This paper proposed architecture for privacy preservation and traceability. The implementation results represent that our method is suitable for large organizations.

## 6. REFERENCES

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

[2] InduArora ,Dr.AnuGupta,"Clouddatabase:A paradigm shift in Databases"IJCI international journal,july 2012.

[3] 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.

[4]T. Mather, S. Kumaraswamy, and S. Latif, Cloud Security and Privacy:An Enterprise Perspective on Risks and Compliance. Sebastopol,CA, USA: O'Reilly Media, Inc., 2009.

[5] Lucca Ferretti,FabioPierazzi,MichelColajani and Micro Marchetti "Performance and cost evaluation ofan adaptive encryption architecture for cloud databases"IEEE transactions on cloud computing,vol 2,no.2,April-June 2014

[6] Boyang Wang, Baochun Li and Hui Li, Oruta: Privacy-Preserving Public Auditing for Shared Data in the Cloud, IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. 2, NO. 1, JANUARY-MARCH 2014, pp. 43-57

[7] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner,Z. Peterson, and D. Song, "Provable Data Possession atUntrusted Stores," Proc. 14th ACM Conf. Computer and Comm.Security (CCS '07), pp. 598-610, 2007.

[8] H. Shacham and B. Waters, "Compact Proofs of Retrievability,"Proc. 14th Int'l Conf. Theory and Application of Cryptology and InformationSecurity: Advances in Cryptology (ASIACRYPT '08), pp. 90-107, 2008.

[9] C. Erway, A. Kupcu, C. Papamanthou, and R. Tamassia,"Dynamic Provable Data Possession," Proc. 16th ACM Conf. Computerand Comm. Security (CCS'09), pp. 213-222, 2009.

[10] Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou, "Enabling PublicVerifiability and Data Dynamic for Storage Security in CloudComputing," Proc. 14th European Conf. Research in Computer Security(ESORICS'09), pp. 355-370, 2009.

[11] C. Wang, Q. Wang, K. Ren, and W. Lou, "Ensuring Data StorageSecurity in Cloud Computing," Proc. 17th Int'l Workshop Quality ofService (IWQoS'09), pp. 1-9, 2009.

[12] B. Chen, R. Curtmola, G. Ateniese, and R. Burns, "Remote DataChecking for Network Coding-Based Distributed Storage Systems,"Proc. ACM Workshop Cloud Computing Security Workshop(CCSW'10), pp. 31-42, 2010.

[13] Y. Zhu, H. Wang, Z. Hu, G.-J. Ahn, H. Hu, and S.S Yau,"Dynamic Audit Services for Integrity Verification of OutsourcedStorages in Clouds," Proc. ACM Symp. Applied Computing(SAC'11), pp. 1550-1557, 2011.

[14] N. Cao, S. Yu, Z. Yang, W. Lou, and Y.T. Hou, "LT Codes-BasedSecure and Reliable Cloud Storage Service," Proc. IEEE INFOCOM,2012.

[15] C. Wang, S.S. Chow, Q. Wang, K. Ren, and W. Lou, "Privacy-PreservingPublic Auditing for Secure Cloud Storage," IEEE Trans.Computers, vol. 62, no. 2, pp. 362-375, Feb. 2013.