A System to Sustain Identity Secrecy of a user on Shared Data in the Cloud

Madhura J
Assistant Professor,
Department of CSE
DSCE, Bengaluru, India,

Manish Bhardwaj
B.E, M. Tech
Bengaluru, India

***Abstract*— Data can be stored in cloud and also shared among multiple users. But the correctness of data present in the cloud may be in doubt as there exists human errors and also some hardware/software failures. Cloud data can be audited by both data owners and public verifiers. Many mechanisms exist through which data can be audited efficiently for checking reliability without retrieving the entire data from the cloud server. But, public auditing for checking the integrity of shared data will predictably reveal some private information about the identity to public verifiers. Therefore a unique mechanism to preserve identity privacy which supports public auditing for the shared data that is stored in cloud is proposed. Here ring signatures that are required to compute verification metadata for auditing the integrity of shared data are exploited. With this mechanism, the shared data integrity can be efficiently verified without retrieving entire file by not disclosing the identity of the signer on each block to the public verifier. Instead of verifying one by one , the mechanism will also perform multiple auditing tasks simultaneously.**

Keywords*— Cloud computing, privacy preserving, public auditing ,shared data.

I. INTRODUCTION

Cloud computing is a group of IT possessions that are made available to a customer over a set of connections on a leased basis and with the ability to balance up or down their service necessities . These services that are delivered are called as Software as a Service. In SaaS, the computing resources are given in the form of service instead of product to the customers through the internet. Public cloud is the one in which the cloud is delivered to the common public in the form of pay-as-you-go. Utility computing refers to the service that is being sold. Cloud computing an aggregate of Software as a Service as well as utility computing. It do not comprise of private cloud in it.

The standard feature of cloud storage offering is data sharing, including Drop box and GoogleDocs. The data can be shared with many users in a group. Third party auditor can be introduced to perform public auditing for protecting correctness of data that is placed in cloud. Auditing benefits are recommended by the TPA with additional authoritative computation normal users. The most important feature that motivated cloud storage is data sharing among several users. Retrieving the entire data from the cloud and then checking the integrity by verifying signatures for its correctness or hash values of the entire data is what is followed in the traditional approach in order to make sure the reliability of data that is maintained in the cloud.

With the aid of traditional approach the correctness of cloud data can be checked successfully but the efficiency is in doubt. Large size of cloud data is the main reason for the inefficiency. If in case the data that is placed in the cloud is corrupted, then downloading the whole data from the cloud for checking the integrity will waste many communication resources and will cost much of the user computation amount. Hence it is very much needed to ensure security for the whole data that is placed in the cloud. Proper mechanisms should be employed in order to address the above mentioned issue.

Data mining, machine learning and many other uses of cloud data do not exactly need users to download entire cloud data because user’s computation services will be provided directly on large scale data by many cloud service providers such as Amazon that is already there in the cloud. Many mechanisms have been proposed which allows both the data owner and the public verifier to proficiently carry out integrity scrutiny which is known as public auditing for the data that is placed in the cloud. Here, the data that is maintained in the cloud will be divided into number of small blocks. Each and every block is signed autonomously by the owner. During reliability checking, arrangement of all blocks will be repossessed instead of the intact file.

Cloud storage is mainly motivated by one of the most important feature called as data sharing among many users. Hence, ensuring shared data integrity in the cloud is also necessary. In order to verify shared data integrity, the existing mechanisms can also be extended. However, in the case of data that is commonly shared, a privacy issue is introduced. With existing mechanisms the identity details will be leaked to the third party who carries out the public auditing.

For instance, A and B makes a group and works together in which they together share their data file in the cloud. With the help of present methods each and every block is signed individually by the users such as X and Y by separating the entire data into many little blocks. When a block is modified by a user in this shared file, then this user must use the private key that is with them in order to sign the new block. Owing to the modifications introduced by these two unlike users the modified blocks are signed by different users. Then, an appropriate public key for each block must be choosen properly by the public verifier to correctly validate the integrity of entire data stored in the cloud.

As a result, the uniqueness of the signer on individual blocks will be known by the public verifier under public key infrastructure(PKI) through which the exclusive binding between an individuality and a public key by means of digital certificates can be recognized. If identity privacy is not preserved during public auditing on the data that is shared in the cloud, then many important confidential data or information(e.g., particularly who is the fundamental user in the group or which is the unique block in the collective data) will be leaked to the public verifiers. Specifically, as shown in Fig.1, [11] when auditing task is carried out many times, the public auditor can first get to know that Alice may be a prime important user in the group since Alice has signed majority of the blocks in the shared file. Additionally, the public auditor can also easily come to a conclusion that the eighth block may contain very important and confidential data (e.g., a final tender in an public sale), because the two different users have modified this block very frequently. Hence to guard these secret information, it is very necessary and highly important to maintain identity privacy from the public auditor during the task of public auditing.

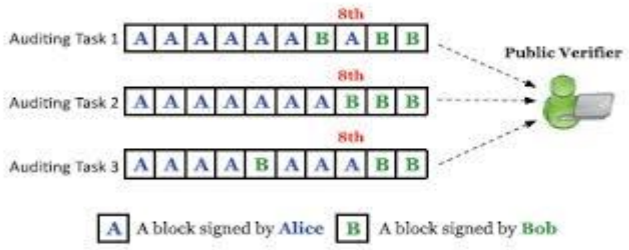

Fig. 1: A and B share a data file in the cloud, and public verifier audits shared data integrity with existing mechanisms

In order to overcome the above mentioned privacy concern introduced on shared data, a contemporary mechanism for preserving identity privacy is proposed which is known as ORUTA. Ring signatures are specifically made use to construct homomorphic authenticators in Oruta, so that the correctness of shared data can be confirmed by a public verifier without retrieving the whole data while the distinctiveness privacy of the signer on each block in data that is shared in the cloud is reserved very confidential to the public verifier. Further, the mechanism is also extended to sustain batch auditing, with the help of which several auditing tasks can be performed concurrently and hence the efficiency of verification for multiple auditing tasks can also be increased.

II. LITERATURE SURVEY

Cloud is a large assembly of systems which are coupled to each other. It is a modification in the manner information is gathered and the application is handled. Cloud computing is a common collection of computing possessions that are configurable, network admittance which are on-demand and looked after by the service provider. Cloud computing is a extensive vision of processing service that facilitates the services to be shared over the internet. Security is the key disadvantage. Many software industries mainly use cloud computing. A unique security arrangement is made by the companies in order to provide security. Since the data stored in the cloud is obtainable by everyone, security is not made sure.

Cloud Computing means providing services, for instance applications over the internet and the hardware and systems software that are available in the datacenters which impart those services. The services that are offered are referred to as Software as a Service (SaaS). "The hardware and system software stored in the datacenter is called as Cloud". Public Cloud is the one which is made available to the common public in the form of pay-as-you-go. The services that are being available for use are called Utility Computing. The internal datacenters of an enterprise or association which are not provided for the unanimous public is called as private cloud. Hence, the total of SaaS as well as Utility Computing is referred to as Cloud computing, whereas it does not comprise of private clouds in it. People can just be the client or contributors of SaaS and Utility Computing. The concentration is mainly on cloud suppliers and cloud users . SaaS users gain more attention.

There are three categories of cloud namely public cloud, private cloud and hybrid cloud. The cloud in which the computing infrastructure is offered by the cloud vendor is identified as public cloud. The user will have no idea of where the infrastructure is placed and will also not have the control on it. Cloud infrastructure is shared among many enterprises. Private cloud is committed to a particular enterprise and not accessible by many enterprises. When compared to public clouds, private cloud is more secure and costly. Hybrid clouds are the grouping of both public and private clouds. It is also called as "cloud bursting". Community cloud means sharing of infrastructure among the same enterprises of same community[1].

When huge amount of data is to be stored by any enterprise then cost will increase. Users should be provided with a facility to utilize the cloud storage as if it is confined without thinking about the necessity to make sure the exactness of data that is stored in cloud. Carrying out public auditing for the data that is placed in the cloud is of very much importance. Users can present it to a third party assessor to test the accuracy of data, so that they can be tension free. An effective third party auditor should be introduced for carrying out the auditing process which should not pose any new problems regarding privacy of data stored in the cloud by the user and should not fetch any new added inconvenience for the user. Hence a confined cloud storage system which supports confidentiality preserving public auditing is presented. Additionally it is extended to enable the TPA to carry out auditing for multiple users at the same time and efficiently[2].

The user who has placed the data can be provided with an opportunity to ensure that whatever data they have stored in some untrusted server is not modified or corrupted by anyone and also that the server always maintains the original data without recovering the whole data with the help of a model known as provable data possession (PDP). This representation generates probabilistic confirmation of ownership by a variety of arbitrary series of blocks from the server, which diminishes I/O costs to a large extent. The verification is sustained by the client and a constant quantity of metadata is used to authenticate the proof. The challenge/response set of rules broadcasts a small, steady amount of data that minimizes network communication. As a result, the PDP model for distant data checking supports huge data collection in generally distributed storage systems. Two provably secure PDP methods that are further proficient than previous solutions, even when evaluated with systems that enable weaker guarantees is offered[3].

The provision provided for data maintenance along with integrity through the most used types of data operations like modifications in the blocks, inserting more data, deleting existing data is very important regarding realism. This is because the services provided through cloud is not confined only to document and support data. The main disadvantage associated with prior mechanisms regarding the data place at a distant place is that both public auditing and dynamic operations carried on data are not sustained. In order to overcome the above mentioned issues, the complexity and security related problems must be addressed initially and then the issue related to dynamic operations should be resolved. Later, an more secure scheme that aggregates the mentioned features should be designed[4].

The important issue introduced in the cloud is security, which is in concern with the data storing process in the untrusted servers of the cloud and the networks in which the resources are shared. In the dynamic provable data possession (PDP) scheme, the user deals with the data prior itself and later transmits it to some untrusted server for storage. During this process it preserves a very little amount of metadata within it. After this the user requests the server to check whether the data stored by them is changed or modified by anyone. The user desires to carry out this task by the server without downloading the whole data. The existing PDP model was applied only for the static or constant data whereas the dynamic mechanism is done for the data that changes more often also. The PDP model is extended for dynamic data also because to sustain any update to the already stored data if in case it is required[5].

The process of summing up different digital signatures is carried out by aggregate signature method. If there are many signatures present on numerous messages that are signed by variety of users, then it can be grouped together into a single signature of short length. The obtained short message is only enough to indicate to the auditor that many users did sign the particular original messages i.e., user j signed the message Mj for j=1,..,n. Hence an effective aggregate signature from a recently obtained fresh short signature which is based on bilinear maps by Boneh, Lynn, and Shacham is obtained. Importance of aggregate signatures are to minimize the volume of certificate chain by grouping up all the signatures in the chain into one and for reducing message extent in secure routing protocols such as SBGP. It is also showed that aggregate signatures gives scope for verifiably encrypted signatures. Such signatures permits the verifier to examine that a given ciphertext C is the encryption of a signature on a previously stated message M. Verifiably encrypted signatures are exploited in contract-signing protocols[6].

Cloud storage services permit the clients to outsource their data in the cloud storage servers and recover them whenever required. This avoids the cost of building and maintaining their own data store. But the users need to provide privacy for the data and also needs to be able to search it without losing privacy. To keep user’s data confidential against untrusted cloud service provider or outsiders, the normal way is to apply cryptographic primitives by giving the secret keys only to authorized users. An efficient, secure and privacy preserving keyword search scheme which supports multiple users with low computation cost and flexible key management is presented[7].

For proficient data dynamics, the present proof of gathering data representations are improved by controlling the typical method for block tag verification. This method is accomplished to be protected not in favor of an untrusted server. It is also confidential opposed to third party verifiers. All the scrutiny of the proposed scheme has very superior competence in the outlook of communication, calculation and storage overheads. In the existing creation, data parallel dynamics can be sustain by using block level dynamics. Whenever there is a change in the portion of data, the subsequent blocks and tags are revised.

A POR is a technique in which it can be viewed as a form of cryptographic proof of knowledge (POK). It is majorly projected to deal with a large file E. POR consists of a set of rules that are designed in which the communication expenses, total number of memory log on given for the users and storage provisions of the user or verifier are measured as small constraints that are basically independent of the length of E. Additionally to this, a new accurate POR constructions, execution deliberations and optimizations that are allowed on earlier explored and associated schemes are presented.

Here it is not essential for a verifier to have the detailed awareness regarding the contents of the file E like POK. PORs give rise to a inventive and remarkable security definition whose formulation is another contribution of the attempt made. PORs is viewed as "an significant tool for semi-trusted online documentation". Previous cryptographic method enables users to ensure the privacy and reliability of files they recover. It is also not recommended for clients to check the data in order to delete or make changes in the file before the recovery process. The main goal of a POR is to achieve these checks without clients having to download the files themselves. A POR can also present quality-of-service guarantees, i.e., show that a file is retrievable within a certain time bound[8].

The environments with strong bandwidth constraints require Short digital signatures. For example, users are often asked to present key signature that is given on the CD label for the product registration. When a individual said to create a digital signature, the least possible signature with shortest length is preferred. Similarly, due to space limitation, during the process of generating the stamp for a post the shortest signatures are created in the form of a bar code. The concept of group signatures that was introduced by Chaum and van Heyst present vagueness for the signers. It is not mandatory that a particular user should only sign the messages. Any member of the group can carry out this task, but the resulting signature maintains the identity of the particular signer very secretly. When auditing process is given to some third person then they may try to know the signature or undo its confidentiality using some meticulous trapdoor. To maintain security during revocation process, some systems provide a way not to transfer the signing power of a disabled user to any of the present users. At present, the most proficient constructions are based on the Strong-RSA theory initiated by Baric and Pfitzman.

In the past two years various projects were proposed that all made use of group signature concepts. Among many proposed projects the first one was the Trusted Computing effort which enabled a Desktop PC to confirm to a remote party regarding what software it is executing through a process called as attestation. In case of privacy-preserving verification, group signatures are considered to be the best. Perhaps an even more suitable project is the Vehicle Safety Communications (VSC) system from the Department of Transportation in the U.S. The system inserts short-range transmitters in the cars; these transmit status information to other cars in close immediacy. For example, if a car executes an emergency brake, all cars in its vicinity are alerted[9].

III. PROPOSED SYSTEM

A. SYSTEM MODEL

The system model comprises of three parties namely the cloud server, a set of users and a public verifier which is illustrated in FIG 2, [11]. The original user and a number of group users are the two category of users in the group . The original user primarily creates shared data in the cloud which is later shared among the group users. Every member of the group is permitted to access and revise shared data. The data that is shared along with its verification metadata (i.e., signatures) are both saved in the cloud server. The third party auditor who is a public verifier providing proficient data auditing services or a data user outside the group willing to utilize shared data will be able to publicly verify the integrity of shared data that is stored in the cloud server. When a request is made by the public verifier to check the integrity of shared data, it first sends an auditing challenge to the cloud server. After getting the auditing challenge, the cloud server responds to the public verifier with an auditing proof of the ownership of shared data. Later, this public verifier checks the correctness of the complete data by checking and ensuring the correctness of the auditing proof. Effectively, between a public verifier and the cloud server the process of public auditing is very similar to challenge and response protocol.

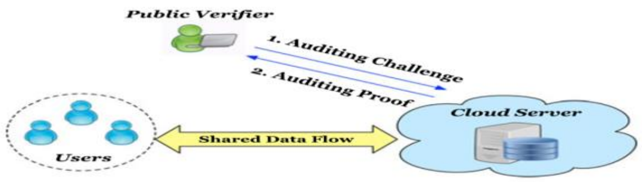

Fig. 2 System model

B. ALGORITHMS

1. RSA Algorithm

RSA stands for Ron Rivest, Adi Shamir and Leonard Adleman. RSA is a public key cryptosystem which is used for secure transmission of data. It contains two types of key namely encryption key and decryption key. Encryption key is public whereas decryption key is private and kept secret. The steps followed for generating a key and for carrying out encryption and decryption is given below

- For the generation of key
  1. Find out two large prime numbers, p and q
  2. Estimate n = pq
  3. Let m = (p-1)(q-1)
  4. Prefer a very small number *e* which is a co-prime to m
  5. Find d, such that de % m = 1
  6. Assign *e* and *n* as the public key.
  7. Remaining d and n are considered as secret key.
- For Encryption
  $C = Pe \ \% \ n$
- For Decryption
  $P = Cd \ \% \ n$

1. *MD5 Algorithm*

The MD5 message-digest algorithm is a widely used cryptographic hash function which produces a 128-bit (16-byte) hash value, normally expressed in text format as a 32 digit hexadecimal number. MD5 has been utilized in a wide variety of cryptographic applications, and is also frequently used to confirm data correctness. The input message is busted up into chunks of 512-bit blocks and the message is padded so that its length is divisible by 512. The padding works as follows: First a single bit, 1 is appended to the end of the message. This is followed by as many zeros as are required to bring the length of the message up to 64 bits fewer than a multiple of 512. The remaining bits are filled up with a 64-bit integer representing the length of the original message, in bits. The MD5 algorithm uses 4 state variables, each of which is a 32 bit integer. These variables are sliced and diced and are the message digest. The major part of the algorithm uses four functions. Those functions are as follows:

$F(X,Y,Z) = (X \& Y) \mid (\sim(X) \& Z)$

$G(X,Y,Z) = (X \& Z) \mid (Y \& \sim(Z))$

$H(X,Y,Z) = X \wedge Y \wedge Z$

$I(X,Y,Z) = Y \wedge (X \mid \sim(Z))$ .The above listed functions that makes use of the state variables and the message as input, are utilized to change the state variables from their initial state into message digest. For each 512 bits of the message, the steps are repeated.

C. WORKING PROCESS

The Working process of the system is indicated in Fig. 3, [11] where the application users will place their data in the cloud. These files will be stored in the database of the cloud service providers. A management server will be placed in the cloud. When a user wants the file to be audited, the request will be sent to the third party auditor. The granted application creates a digital signature1 using MD5 for that particular file and stores. It sends the auditing request to the third party auditor of the file sent by the user. The third party auditor requests the file to be audited from the cloud. Another digital signature2 for that file will be generated and stored in the management server. Instead of sending the file requested by the auditor, the digital signature2 which is generated will be sent. The auditor compares both digital signatures and sends the status report of the file. The status report will be sent as file verified and no modification occurred if both the digital signatures match. If there is a mismatch in the signatures then report will be sent as file verified, modifications occurred, please upload a fresh file.

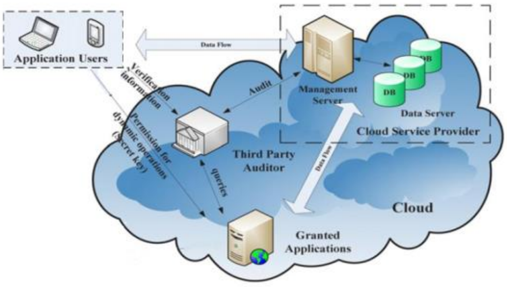

Fig. 3: System Architecture

IV. CONCLUSION AND FUTURE ENHANCEMENT

ORUTA is used as a new mechanism for the purpose of preserving privacy during public auditing in which ring signatures are used to create homomorphic authenticators so that a third party verifier can successfully confirm the exactness of data shared in the cloud without repossessing the complete data from there. Here, individuality of the signer on each block in shared data is kept confidential from the public verifier. The important inferences obtained are as follows

- The individuality of the user who has signed on the data block will not be disclosed.
- The message contents will be kept very private from the third party auditor.
- Dynamic data can also be verified successfully.
- Multiple auditing tasks can be performed concurrently in which the effectiveness will be improved.

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.H. Katz, A. Konwinski, G. Lee, D.A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “A View of Cloud Computing,” Comm. ACM, vol. 53, no. 4, pp. 50-58, Apr. 2010.

[2] C. Wang, Q. Wang, K. Ren, and W. Lou, “Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing,” Proc. IEEE INFOCOM, pp. 525-533, 2010.

[3] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, “Provable Data Possession at Untrusted Stores,” Proc. 14th ACM Conf. Computer and Comm. Security (CCS ’07), pp. 598-610, 2007.

[4] Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou, “Enabling Public Verifiability and Data Dynamic for Storage Security in Cloud Computing,” Proc. 14th European Conf. Research in Computer Security (ESORICS’09), pp. 355-370, 2009.

[5] C. Erway, A. Kupcu, C. Papamanthou, and R. Tamassia, “Dynamic Provable Data Possession,” Proc. 16th ACM Conf. Computer and Comm. Security (CCS’09), pp. 213-222, 2009.

[6] D. Boneh, C. Gentry, B. Lynn, and H. Shacham, “Aggregate and Verifiably Encrypted Signatures from Bilinear Maps,” Proc. 22nd Int’l Conf. Theory and Applications of Cryptographic Techniques: Advances in Cryptology (EUROCRYPT’03), pp. 416-432, 2003.

[7] Y. Zhu, H. Wang, Z. Hu, G.-J. Ahn, H. Hu, and S.S Yau, “Dynamic Audit Services for Integrity Verification of Outsourced Storages in Clouds,” Proc. ACM Symp. Applied Computing (SAC’11), pp. 1550-1557, 2011.

[8] A.Juels and B.S. Kaliski, "PORs: Proofs of Retrievability for Large Files," in Proceedings of ACM CCS'07, 2007, pp. 584-597.

[9] D. Cash, A. Kupcu, and D. Wichs, "Dynamic Proofs of Retrievability via Oblivious RAM," in Proceedings of EUROCRYPT 2013, pp. 279-295.

[10] D. Boneh, B. Lynn, and H. Shacham, “Short Signatures from the Weil Pairing,” Proc. Seventh Int’l Conf. Theory and Application of Cryptology and Information Security: Advances in Cryptology (ASIACRYPT’01), pp. 514-532, 2001.

[11] B. Wang, B. Li, and H. Li, “Oruta: Privacy-Preserving Public Auditing for Shared Data in the Cloud,” IEEE Transactions on cloud computing, vol 2, no. 1, March 2014.