Secure and Efficient Audit Service for Data Integrity in Cloud Storage

T. Poongothai[1],
[1]Professor,
Department of Computer Science and Engineering.
K.S.R. College of Engineering,
Tiruchengode,India.

S. F. Abdul Khadar[2], R. Arunkumar[3],
J. Dinesh[4], M. Hariharan[5]
[2,3,4,5] UG Students,
Department of Computer Science and Engineering.
K.S.R. College of Engineering,
Tiruchengode,India.

Abstract **-Cloud-based outsourced storage relieves the client's burden for storage management and maintenance by providing a comparably low-cost, scalable, location-independent platform. Cloud computing is an emergent paradigm to provide reliable and resilient infrastructure enabling the users (data owners) to store their data and the data consumers (users) can access the data from cloud servers. However, the fact that clients no longer have physical possession of data indicates that they are facing a potentially formidable risk for missing or corrupted data. The Cloud Storage Service (CSS) relieves the burden of storage management and maintenance. To avoid the securityrisks, audit services are critical to ensure the integrity and availability of outsourced data and to achieve digital forensics and credibility on cloud computing. Provable Data Possession (PDP), which is a cryptographic technique for verifying the integrity of data without retrieving it at an untrusted server, can be used to realize audit services. The profiting from the interactive zero-knowledge proof system, address the construction of an interactive PDP protocol to prevent the fraudulence of proven (soundness property) and the leakage of verified data (zero-knowledge property). One fundamental aspect of this paradigm shifting is that data are being centralized and outsourced into clouds. Therefore, it isnecessary for cloud service providers to offer an efficient audit serviceto check the integrity and availability of the stored data.Traditional cryptographic technologies for data integrity and availability, based on hash functions and signature scheme, cannot work on theoutsourced data without a local copy of data.**

Keywords:- Integrity verification, Privacy preserving, Dynamic auditing, Provable DataPossession.

1 .INTRODUCTION

Cloud computing is regarded as such a computing paradigm in which resources in the computing infrastructure are provided as services over the Internet. The cloud computing benefits individual users and enterprises with convenient access, increased operational efficiencies and rich storage resources by combining a set of existing and new techniques from research areas such as service-oriented architectures and virtualization [1]. Although the great benefits brought by cloud computing are exciting for users, security problems may somehow impede its quick development. Currently, more and more users would outsource their data to Cloud Service Provider(CSP) for sharing. However, the CSP which deprives data owners' direct control over their data is assumed to be honest-but-curious, that may prompt security concerns. These security matters existing in public cloud motivate the requirement to appropriately keep data confidential. Several schemes exploiting cryptographic mechanisms to settle the security problems have been proposed. In order to guarantee secure data group sharing, identity-based broadcast encryption scheme is employed in public cloud. The data owners could broadcast their encrypted data to a group of receivers at one time and the public key of the user can be regarded as email, unique id and username. Hence, by using an identity, data owner can share data with other group users in a convenient and secure manner. Attribute-based encryption (ABE) is one of new cryptographic mechanisms used in cloud to reach flexible and fine-grained secure data group sharing. Especially, cipher text-policy ABE allows data owners to encrypt data with an access policy such that only users whose attributes satisfy the access policy can decrypt the data. Privacy preserving is essential to prevent TPA to infer the data using the cloud server's response while auditing. However, the schemes do not achieve privacy preserving requirement. Though data dynamics is an important feature to facilitate the data owners to insert, modify, and delete on a particular block of data, without changing the meta-data of other blocks, the techniques proposed in do not achieve data dynamics requirement. Meanwhile, the schemes like could not achieve batch auditing requirement which ensures that TPA should be capable enough to deal with the multiple numbers of simultaneous verification requests from different DUs. This property is to save computation and communication cost between CSP and TPA. Unfortunately, the schemes, use pairing based cryptographic operations which are intensive computation and need more time.

2. RELATED WORK

There have been numerous works on secure data group sharing and dissemination in public cloud based on various cryptographic primitives such as broadcast encryption and attribute based encryptiondesigned a solution that ensures privacy preserving data sharing based on the role-based access control and cryptographic capabilities of client's browser. Based on orderpreserving encryption and homomorphism encryption to guarantee data confidentiality of database in public cloud.

Proposed a secure data group sharing scheme based on IBBE algorithm, in which data owner can broadcast encrypted data to a group of users at the same time. In order to achieve data collaboration and dissemination, this scheme adopted the PRE technique to allow an authorized proxy to convert an IBBE ciphertext into an identity-based encryption (IBE) ciphertext. Hence, the intended receiver can decrypt the IBE ciphertext. However, this scheme only allows the re-encryption procedure to be executed in an all-or-nothing manner, which means the proxy can either re-encrypt all the initial ciphertexts or none of them. The ciphertext-policy scheme could allow users to generate a re-encryption key associated with a condition and only the encrypted data meeting the condition can be re-encrypted. A conditional identity-based broadcast scheme to achieve secure data group dissemination in cloud email. The CIBPRE scheme adopted IBBE technique to allow a sender to encrypt a message to a group of receivers by specifying the receivers' identities, and the sender can delegate a re-encryption key to a proxy so that he can convert the initial ciphertext into a new one to a new group of intended receivers.

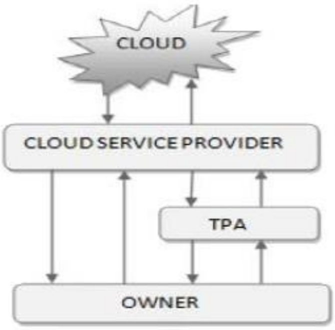


Fig 1: Over all process of the system

2.4 DRAWBACKS STATEMENT

Remote Data Possession Checking (RDPC) is an effective technique to ensure the integrity for data files stored on CSS. RDPC supplies a method for data owner to efficiently verify whether cloud service provider faithfully stores the original files without retrieving it. In RDPC, the data owner is able to challenge the CSS on the integrity for the target file. The CSS can generate proofs to prove that it keeps the complete and uncorrupted data. The fundamental requirement is that the data owner can perform the verification of file integrity without accessing the complete original file. Moreover, the protocol must resist the malicious server which attempts to verify the data integrity without accessing the complete and uncorrupted data. Another desired requirement is that dynamic data operations should be supported by the protocol. In general, the data owner may append, insert, delete or modify the file blocks as needed. Besides, the computing complexity and communication overhead of the protocol should be taken into account for real applications. The present of a novel efficient RDPC scheme with data dynamics. The basic scheme utilizes homomorphic hash function technique, in which the hash value of the sum for two blocks is equal to the product for two hash values of the corresponding blocks. A linear table called ORT(Operational Readiness Test) to record data operations for supporting data dynamics such as block modification, block insertion and block deletion. To improve the efficiency for accessing ORT, we make use of doubly linked list and array to present an optimized implementation of ORT which reduces the cost to nearly constant level.The presented scheme is secure against forgery attack, replay attack and replace attack based on atypical security model.

3. RECENT METHODS

Public Provable Data Possession (PDP) has been used, which is a cryptographic technique for verifying the integrity of data without retrieving it at an untrusted server; can be used to realize audit services. It we random mask technique to achieve a privacy-preserving public auditing system for cloud data storage security while keeping all above requirements in mind to support efficient Handling of multiple auditing tasks and further explore the technique of bilinear aggregate signature to extend our main result into a multiuser setting, where TPA can perform multiple auditing tasks simultaneously.

Extensive security and performance analysis shows the proposed schemes are provably secure and highly efficient and also show how to extent our main scheme to support batch auditing for TPA upon delegations from multi-users.

4. PROPOSED WORK

Public Provable Data Possession (PDP) has been used, which is a cryptographic technique for verifying the integrity of data without retrieving it at an untrusted server; can be used to realize audit services. It we random mask technique to achieve a privacy-preserving public auditing system for cloud data storage security while keeping all above requirements in mind to support efficient Handling of multiple auditing tasks and further explore the technique of bilinear aggregate signature to extend our main result into a multiuser setting, where TPA can perform multiple auditing tasks simultaneously. Extensive security and performance analysis shows the proposed schemes are provably secure and highly efficient and also show how to extent our main scheme to support batch auditing for TPA upon delegations from multi-users.

One of the important concerns that need to be addressed is to assure the customer of the integrity i.e. correctness of his data in the cloud. As the data is physically not accessible to the user the cloud should provide a way for the user to check if the integrity of his data is maintained or is compromised. This proof can be agreed upon by both the cloud and the customer and can be incorporated in the Service Level Agreement (SLA). It is important to note that our proof of data integrity protocol just checks the integrity of data i.e. if the data has been illegally modified or deleted. To propose a data correctness scheme which involves the encryption of the few bits of data per data block thus reducing the computational overhead on the clients. The client storage overhead is also minimized as it does not store any data with it and it reduces bandwidth requirements. In our data integrity protocol the TPA needs to store only a single cryptographic key irrespective of the size of the data file F and two functions which generate a random sequence. The TPA does not store any data with it. The TPA before storing the file at the archive,pre-processes the file and appends some Meta data to the file and stores at the archive. At the time of verification the TPA uses this meta data to verify the integrity of the data. It is important to note that our proof of data integrity protocol just checks the integrity of data. But the data can be stored, that is duplicated at redundant data centres to prevent the data loss from natural calamities. If the data has to be modified which involves updating, insertion and deletion of data at the client side, it requires an additional encryption of fewer data bits. So this scheme supports dynamic behaviour of data.

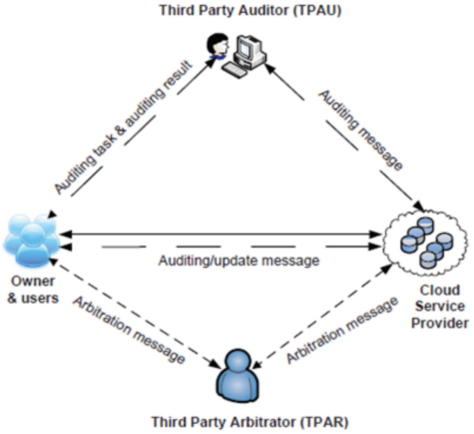


FIG 2: Architecture of the Proposed System

5. IMPLEMENTATION

5.1 DATA STORAGE

Data outsourcing to cloud storage servers is raising trend among many firms and users owing to its economic advantages. This essentially means that the owner (client) of the data moves its data to a third party cloud storage server which is supposed to - presumably for a fee - faithfully store the data with it and provide it back to the owner whenever required. As data generation is far outpacing data storage it proves costly for small firms to frequently update their hardware whenever additional data is created. Also maintaining the storages can be a difficult task. Storage outsourcing of data to a cloud storage helps such firms by reducing the costs of storage, maintenance and personnel. It can also assure a reliable storage of important data by keeping multiple copies of the data thereby reducing the chance of losing data.

5.2 DATA PROOFS

The developing TPA can proofs for data possession at entrusted cloud storage servers we are often limited by the resources at the cloud server as well as at the client. Given that the data sizes are large and are stored at remote servers, accessing the entire file can be expensive in I/O costs to the storage server. Also transmitting the file across the network to the client can consume heavy bandwidths. Since growth in storage capacity has far outpaced the growth in data access as well as network bandwidth, accessing and transmitting the data. Furthermore, the I/O to establish the data proof interferes with the on-demand bandwidth of the server used for normal storage and retrieving purpose. The problem is further complicated by the fact that the owner of the data.

5.3 DYNAMIC PROVISIONING

Allows for the provision of services based on current demand requirements. This is done automatically using software automation, enabling the expansion and contraction of service capability, as needed. This dynamic scaling needs to be done while maintaining high levels of reliability and security.

5.4 VERIFICATION PHASE

The TPA wants to verify the integrity of the file F. It throws a challenge to the archive and asks it to respond. The challenge and the response are compared and if the result is TRUE the TPA accepts the integrity proof. Else if the result of comparison is FALSE it rejects the integrity proof. Suppose the verifier wishes to check the integrity of ith block the TPA challenges the cloud storage server by specifying the block number i and a bit number j generated by using the function g which only the TPA knows. The TPA also specifies the position at which the meta data corresponding to the block i is appended. Hence the cloud storage server is required to send the bits for verification by the client. The corresponding data in meta data is compared with the the cloud. Any mismatch between the two would mean a loss of the integrity of the client's data at the cloud storage.

5.5 GRANULARITY OF INTEGRITY

Data integrity/authentication can be provided at different levels of granularity. This is clearly impractical as it requires transferring large amounts of data to the client. Hence, do not consider this to be a viable approach. On the other hand, computing integrity checks at the level of individual attribute values yields a very large number of signatures which is very expensive for the signer (owner) in terms of computation as well as for the server in terms of storage. The optimal choice is to provide integrity at the record level. This enables the server to return –in response to a query any set of matching records along with their respective integrity checks. Of course, computing integrity checks over the entire record, as opposed to individual attributes, implies that the smallest unit of data returned as a query reply is an entire record, even when the querying client is only interested in a single.

5.6 DYNAMIC OPERATIONS

A straightforward way is to let the arbitrator keep a copy of the index switcher. Since the change of the index switcher is caused by dynamic operations, the client can send necessary update information to the TPAR for each update operation. With these information, the arbitrator could re-construct the latest version of the index switcher, whose correctness decides the validity of later arbitration.

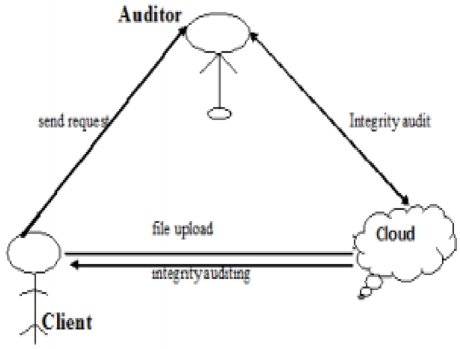
However, such a solution costs storage at the arbitrator side and needs the arbitrator to be involved in each update operation. Ideally, we want the TPA only undertake the role of an arbitrator who involves only at dispute settlement, and maintains a constant storage for state information, i.e., public keys of the client and the CSP.

6. RESULTS AND DISCUSSION

DATA GENERATION

The first step of verifying data integrity is to generate data, including business real data and control data. In the rest of the document, refer to this step as the initializing step. This step starts with determining which data items are sensitive, which discussed previously. This part also mentions which data items should be defined as sensitive. After determining the sensitive data items, should continue the process by adding one into the data dictionary table. The presented structure of the data and the information that keep in this table; this table is fundamentally important for checking integration of real data.

DATA VERIFICATION

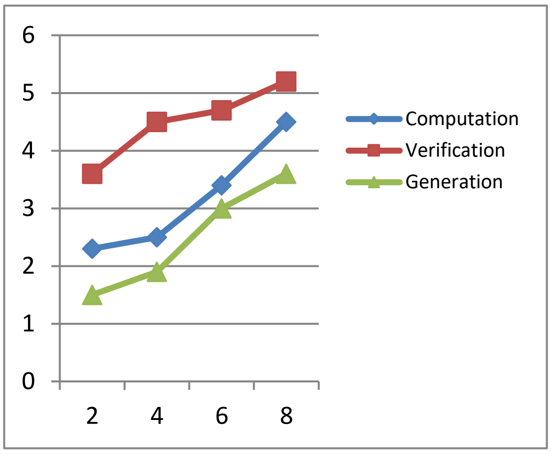


Public verification techniques allow the users to outsource their data into the cloud and consistency Of the data is checked by a trusted third party called auditor. Objective of the public verification scheme is to avoid external adversary attacks on the data outsourced by the owner. In proposed a scheme to remove burden on the user for checking the data integrity by assigning this task to a third party called auditor. It is assumed that auditor is trust worthy and honest in his task. To securely introduce an effective TPA, the auditing process should bring in no new vulnerabilities toward user data privacy, and introduce no additional online burden to use. Here Homomorphic linear authenticator and random masking techniques are combined to achieve privacy preserving auditing scheme. In proposed protocol, the linear combination of sampled blocks in the server's response is masked with randomness generated by the server. With random masking, the TPA no longer has all the necessary information to build up a correct group of linear equations and therefore cannot derive the user's data content, no matter how many linear combinations of the same set of file blocks can be collected.

TABLE SHOWS THAT THE COMPARISION OF DATA VERIFICATION AND DATA GENERATION

S.No	Computation	Verification	Generation
1	2.3	3.6	1.5
2	2.5	4.5	1.9
3	3.4	4.7	3
4	4.5	5.2	3.6

COMPARISION OF DATA VERIFICATION AND DATA GENERATION



DESCRIPTION

The above chart is used to analyse the data verification ,computation and generation of the file in the server of the cloud. Data can be verified at the time of the admin upload file to the server. TPA is used to generate the file verification after that final data file are allowed to the server. All the process the verification level is fast and completed the process easily.

7. CONCLUSIONS

In the paper worked to facilitate the client in getting a proof of integrity of the data which he wishes to store in the cloud storage servers with bare minimum costs and efforts. Our scheme was developed to reduce the computational and storage overhead of the client as well as to minimize the computational overhead of the cloud storage server. We also minimized the size of the proof of data integrity so as to reduce the network bandwidth consumption. At the client we only store two functions, the bit generator function g, and the function h which is used for encrypting the data. Hence the storage at the client is very much minimal compared to all other schemes that were developed. Hence this scheme proves advantageous to thin clients like PDAs and mobile phones. The operation of encryption of data generally consumes a large computational power. In our scheme the encrypting process is very much limited to only a fraction of the whole data thereby saving on the computational time of the client. Many of the schemes proposed earlier require the archive to perform tasks that need a lot of computational power to generate the proof of data integrity. But in our scheme the archive just need to fetch and send few bits of data to the client.

REFERENCES:

[1] Ari Juels and Michael Szydlo, "Attribute-Based Encryption: Using Identity-Based Encryption for Access Control", RSA Laboratories Bedford, MA 01730, 17 June 2004.

[2] Yao Zheng, "Privacy-Preserving Personal Health Record System Using Attribute-Based Encryption", WORCESTER POLYTECHNIC INSTITUTE In partial fullfilment of the requirements for the Degree of Master of Science, June 2011.

[3] Meenal Jain and Manoj Singh, "Identity Based and Attribute Based Cryptography: A Survey", International Journal of Engineering, Management & Sciences (IJEMS) ISSN-2348 – 3733, Volume-2, Issue-5, May 2015.

[4] Ms.DipaliPatil, Dr.P.K.Deshmukh, "Data Security in Cloud Using Attribute Based Encryption with Efficient Keyword Search", International Journal of Scientific & Engineering Research, Volume 7, Issue 1, January,2016.

[5] XingbingFu,Zufeng Wu, "Ciphertext Policy Attribute Based Encryption with Immediate Attribute Revocation for Fine-Grained Access Control in Cloud Storage", School of Computer Science and Engineering, University of Electronic Science and Technology of China, No.2006, Xiyuan, IEEE, 2013.

[6] Jinguang Han, Willy Susilo, Yi Mu and Jun Yan, "Privacy-Preserving Decentralized Key-Policy Attribute-Based Encryption", IEEE Transactions on Parallel and Distributed systems, VOL. 23, NO. 11, NOVEMBER 2012.

[7] Ning Cao, Cong Wang, Ming Li, Member, KuiRen and Wenjing Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data", IEEE Transactions on Parallel and Distributed Systems, VOL. 25, NO. 1, JANUARY 2014.

[8] Giuseppe Ateniese, Roberto Di Pietro, Luigi V. Mancini and Gene Tsudik, "Scalable and Efficient Provable Data Possession", 2008.