IJCSN International Journal of Computer Science and Network, Volume 5, Issue 2, April 2016
ISSN (Online) : 2277-5420 www.IJCSN.org
Impact Factor: 1.02

Comparative Study on Preserving Data Integrity for Shared Data in Cloud Storage

[1] Swapnil Patil, [2] Mettu Govind Rao

[1] PG Student, PHCET,
Rasayani, Maharashtra, India

[2] Assistant Professor, PHCET,
Rasayani, Maharashtra, India

Abstract - A novel privacy-preserving mechanism that supports public auditing for shared data stored on cloud. "Ring Signatures" is a technique to compute authentication metadata required to review the accuracy of shared data. This mechanism precise that the identity of the signer on each block in shared data is hidden from public verifiers, who are able to proficiently verify shared data integrity without retrieving the entire file. Beside the mechanism is able to execute multiple auditing tasks simultaneously instead of verifying them one by one. And with data storage and sharing services in the cloud, users can easily transform and share data as a group. To certify that the shared data integrity can be verified publicly, users in a group need to calculate signatures on all the blocks in shared data. Blocks in these shared data are usually signed by different user due to data alterations performed by different user. For security measures, when user is revoked from a group, the blocks which were previously signed by the revoked user must be re-signed by an existing user of a group.

Keywords – *Privacy Preserving, TPA, CSP.*

1. Introduction

Cloud computing has been recommended for the next generation information technology (IT) architecture for enterprises On demand self-service, ubiquitous network access, location independent resource pooling, rapid resource elasticity, usage-based pricing and transference of risk. As a disruptive technology with reflective implications, cloud computing is altering the environment of how businesses use information technology. One fundamental aspect of this model shifting is that data is being centralized or outsourced to the cloud. From users viewpoint IT enterprises storing data remotely to the cloud in a flexible on-demand manner brings smart benefits as relief of the liability for storage management, universal data access with location independence and avoidance of capital expenditure on hardware, software, and personnel maintenances, etc.

While cloud computing makes these advantages more attractive than ever it also brings new and challenging security threats toward user's outsourced data. Since cloud service providers (CSP) are separate administrative entities data outsourcing is actually giving user ultimate control over the destiny of their data. As a result, the accuracy of the data in the cloud is being put at risk due to the following reasons. First of all, although the infrastructures under the cloud are much more powerful and reliable than personal computing devices, they are still facing the broad range of both internal and external threats for data integrity. Examples of outages and security obstacles of significant cloud services appear from time to time. Second, there do exist various incentives for CSP to behave unfaithfully toward the cloud users regarding their outsourced data status.

For examples, CSP might reclaim storage for financial reasons by discarding data that have not been or are rarely accessed or even hide data loss incidents to maintain a status. In short, although outsourcing data of the cloud is economically attractive for long-term and large-scale storage, it does not immediately offer any guarantee on data integrity and availability. This problem, if not properly addressed, may delay the success of cloud architecture. Last but not least, groups are normally dynamic in practice, e.g., new staff participation and current employee revocation in a company. The changes of membership make secure data sharing extremely challenging. On other hand, the unidentified system challenges new granted users to learn the content of data files stored before their participation, because it is

impossible for new granted users to contact with unidentified data owners, and obtain the corresponding decryption keys.

On the other hand, an efficient membership revocation mechanism without updating the secret keys of the remaining users is also desired to minimize the complexity of key management. Several security schemes for data sharing on untrusted servers have been proposed. In these approaches, data owners store the encrypted data files in untrusted storage and distribute the corresponding decryption keys only to authorized users. Thus, unauthorized users as well as storage servers cannot learn the content of the data files because they have no knowledge of the decryption keys.

However, the complexities of user participation and revocation in these schemes are linearly increasing with the number of data owners and the number of revoked users, respectively. By setting a group with a single attribute, in this proposed a secure provenance scheme based on the cipher text-policy attribute-based encryption technique, which allows any member in a group to share data with others. However, the issue of user revocation is not addressed in their scheme. Yu et al presented a scalable and fine-grained data access control scheme in cloud computing based on the key policy attribute based encryption (KP-ABE) technique.[2][3][4]

2. Literature Survey

Privacy is one of the major concerns when sharing data on network for business analysis. Privacy preservation is the most important area in today's computing field. For this purpose, there are different techniques given as Computing is combination of a set of software framework, infrastructure, and middleware services that can allow sharing and selection of resources. Grid computing have different types of parameter such as software framework, infrastructure and middleware service allowing seamless sharing, aggregation and selection of resources across multiple heterogeneous control and administrative domains, etc. In paper "Privacy Protection In Anonymous Computational Grid Services", Service-oriented architecture (SOA) technique implemented based on grid backbone and they performed different functions like stakeholders are service providers, service requestors or service consumers and service brokers. In this research author, used service requestor communicating through Java enabled web browsers. The server or service providers or the service collators had the web page container. When the client browser intends to load with JAVA designated servlet, by using communication and it can be continued through SOAPXML messages. In between multiple service providers to collate the responses to one single colossal query by using dynamic collaboration. They had an architectural model like that they used a collaborative model. Onion routing is also used for security purpose in gird computing. In this method, only encrypted packet gets transferred among intermediate core nodes, the node that having specific distinct attribute serves as the public key. For private key, dynamically generated token before every hop. The main advantage of onion routing, no encrypted message get lost during travelling in network if any intermediate node failed.

In paper "Privacy Preservation by k-Anonymization of Weighted Social Networks", proposed an anonymization technique for weighted graphs, i.e. for social networks. They proposed a method that Provides k-anonymity of nodes against attacks where the adversary has information about the structure of the network, including its edge weights. In this method, group of different nodes with similar and dis-similar sets of neighbors and their connections into super nodes and edges, respectively. They mainly consider prevention of identity disclosure, but they also touch on edge and edge weight disclosure in weighted graphs. The advantage of this method, it had efficiently work in a weighted graph. Whereas drawback of this method, to preserve utility of the graph. In paper "Anonymizing Classification Data for Privacy Preservation" discussed method TDR, which having the large scale of data sets and complex anonymity requirements. Their research objective in this is to evaluate the method, that is, TDR, for preserving the usefulness of classification and the scalability on large data sets. For the evaluation of usefulness, they compared the classifier built from the masked data as well as unmodified data. In past researched work some model the classification metric on the masked table, the optimality of such metrics does not translate into the optimality of classifiers.

According to author knowledge, classification of anonymity on the basis of single dimensional generalization and on the basis of this impact is evaluated. Because of these reasons, there evaluation used the baseline of the unmodified data and the reported results. All experiments on TDR were conducted on an Intel Pentium IV 2.6-GHz PC with 1-Gbyte RAM. The proposed TDR method having many advantages, it should produce comparable accuracy, TDR much more efficient than previously researched and TDR have also found better anonymization solution for classification. In paper "Privacy Preserving in Data Mining Using Hybrid

Approach", author suggested Oruta method, for privacy preserving public auditing mechanism for shared data in the cloud. There work in this research involves three parties: the cloud server, the third party auditor (TPA) and users.[6] There are two types of users in a group: the original user and a number of group users. The original user and group users are both members of the group. Group members are allowed to access modify shared data created by the original user based on access control polices. Shared data and its information are both stored in the cloud server. On behalf of group members, third party auditor had verified the integrity of shared data in the cloud server. This method worked efficiently and securely to verify shared data for a group of users. The third party auditor have able to correctly detected whether they had any corrupted block in shared data. There are many limitations of Oruta, only group generated valid verification information on shared data, it could not identified single user on shared data and also large communication cost.

Table 1. Comparative Study of Different Techniques

Sr.No.	Name of Technique	Advantages	Drawbacks
1	Grid Computing	Preventing both intersection and predecessor attacks as well as protected from eavesdropping and traffic analysis	Updated Periodically
2	k-anonymity	Gives privacy protection and usability of data	Homogeneity-y and background attack
3	TDR	Scalability for large data sets	Support only single dimension a generalization
4	TPA	Work efficiently and securely	Large communicate-on cost
5	Repository System	Larger data distribution capability	difficult for the eavesdroppers

In paper "Oruta: Privacy-Preserving Public Auditing for Shared Data in the Cloud", proposed technique k-degree-diversity model for data publishing in privacy preserving Social network.

They introduced a noise node adding algorithm to construct a new graph from the original graph with the help of constraint of introducing fewer distortions to the original graph. They gave analysis of the theoretical bounds on the number of noise nodes added and their effects on an important graph property. There extensive experimental results demonstrated using noise node adding algorithms could achieved a better result than the previous worked by edge editing only. For generating a k-neighborhood anonymous graph was sort all nodes by their neighborhood graph size in descending order and then recursively adjust two nodes neighborhood graphs to be the same until a k-neighborhood anonymity graph was generated. An unanonymzed node having the smallest degree with the highest priority to be added. The noise node adding strategy should be considered in this step to improve the utility of the published graph. This research having many advantages, it can reduce noise nodes, a unified data together to guarantee the privacy. Whereas drawback of this method data should be overlapped. In paper "Protecting Sensitive Labels in Social Network Data Anonymization", proposed technique trust-based privacy preservation method for data sharing in P2P.

In a P2P system, in which privacy is different from the traditional node anonymity problem and the identities of the participants are known. During data acquirement a peer acted as the proxy server. To get the data through proxy server, requester sends the request first which made it difficult for the eavesdroppers. A privacy measuring method is given to evaluate the proposed research. A recommendation caused the largest change in the trust value is predetermined by the peers. Recursive calculation done by the fading speed and the mapping functions. Least-square-error method are determined the parameters. When a threshold exceeds then the difference between the predicted trust value and the observed value and the algorithm will change the values of the parameters. This makes the algorithm the changes in peer's behavior. The advantages of this method efficiency/accuracy trade-off in trustworthiness assessment. The drawback of this method lack of privacy

In paper "Trust-Based Privacy Preservation for Peer-to-Peer Data Sharing" suggested technique privacy preserving repository to integrate data from various data sharing services. The performance of decryption depends

on these parameter repository only collects the minimum amount of information from data sharing services based on users' integration requests, and data sharing services can restrict our repository to use their shared information only for users' integration requests, but not other purposes. They assume that our repository can access all shared data and focus on how data sharing services shared data for specific data integration requests to prevent our repository from using the shared data for other purposes. In this research they have only focused on matching operations and additive homomorphism encryption schemes there repository could be easily extended to support SUM and AVG aggregate operations. The drawback of this system it cannot worked for large scale data sets and enable there repository to support more types of data integration operations. For future research need to investigate the behavior of our repository when there are conflicts among data sharing services' policies on the shared data.[1]

3. Existing System

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. It transmitting the file across the network to the client can consume heavy bandwidths. The problem is further complicated by the fact that the owner of the data may be a small device, like a PDA (personal digital assist) or a mobile phone, which have limited CPU power, battery power and communication bandwidth.

Several security schemes for data sharing on untrusted servers have been proposed. In these approaches, data owners store the encrypted data files in untrusted storage and distribute the corresponding decryption keys only to authorized users. Thus, unauthorized users as well as storage servers cannot learn the content of the data files because they have no knowledge of the decryption keys. However, the complexities of user participation and revocation in these schemes are linearly increasing with the number of data owners and the number of revoked users, respectively. By setting a group with a single attribute, Lu et al.[5] proposed a secure provenance scheme based on the cipher text-policy attribute-based encryption technique, which allows any member in a group to share data with others. However, the issue of user revocation is not addressed in their scheme. Yu et al.[5] presented a scalable and fine-grained data access control scheme in cloud computing based on the key policy attribute-based encryption (KP-ABE) technique. Unfortunately, the single owner manner hinders the adoption of their scheme into the case, where any user is granted to store and share data.

3.1 Disadvantages of Existing System

- The main drawback of this scheme is the high resource costs it requires for the implementation.
- Also computing hash value for even a moderately large data files can be computationally burdensome for some clients (PDAs, mobile phones, etc).
- Data encryption is large so the disadvantage is small users with limited computational power (PDAs, mobile phones etc.).
- Identity privacy is one of the most significant obstacles for the wide deployment of cloud computing. Without the guarantee of identity privacy, users may be unwilling to join in cloud computing systems because their real identities could be easily disclosed to cloud providers and attackers. On the other hand, unconditional identity privacy may incur the abuse of privacy.
- It is highly recommended that any member in a group should be able to fully enjoy the data storing and sharing services provided by the cloud, which is defined as the multiple-owner manner. Compared with the single-owner manner, where only the group manager can store and modify data in the cloud, the multiple-owner manner is more flexible in practical applications.
- Groups are normally dynamic in practice, e.g., new staff participation and current employee revocation in a company. The changes of membership make secure data sharing extremely difficult.

4. Proposed System

One of the important concerns that need to be addressed is to assure the customer of the integrity i.e. correctness of their data in the cloud. As the data is physically not available to the user the cloud should provide a way for the user to check if the integrity of their data is maintained or is compromised. So using the new scheme which gives a proof of data integrity in the cloud which the customer can employ to check the correctness of his data in the cloud. Evidence can be recognized upon both the cloud and the customer and can be integrated in the Service level agreement (SLA). It is important to note that

proof of data integrity protocol just checks the integrity of data i.e. if the data has been illegally altered or deleted also propose a secure multi-owner data sharing scheme for dynamic groups in the cloud. With the help of group signature and dynamic broadcast encryption techniques any cloud user can anonymously share data with others. It means the storage overhead and encryption computation cost of our scheme are independent with the number of revoked users. As well for scrutinize the security scheme with rigorous proofs, and demonstrate the efficiency scheme in experiments.

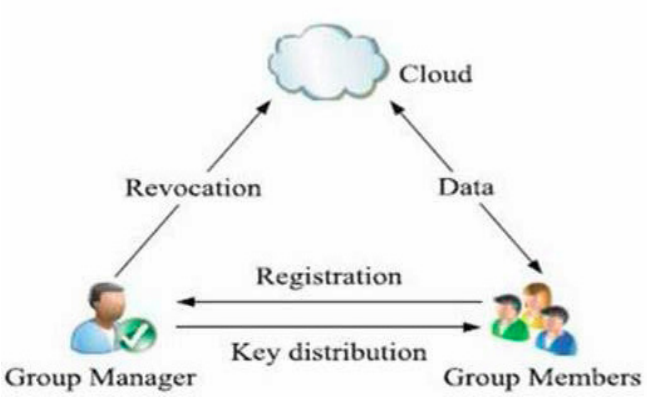

Fig.1 Working of Proposed System

4.1 Advantages of Proposed System

- Apart from decrease in storage costs data outsourcing to the cloud also helps in reducing the maintenance.
- Avoiding local storage of data.
- By decreasing the costs of storage, maintenance and personnel.
- It reduces the chance of misplacing data by hardware failures.
- Any user in the group can store and share data files with others by the cloud.
- The encryption complexity and size of cipher-texts are independent with the number of revoked users in the system.
- User revocation can be achieved without updating the private keys of the remaining users.
- A new user can directly decrypt the files stored in the cloud before his participation

5. Implementation Methodologies

5.1 Cloud Module

In this module, local Cloud platform provide an abundant storage services so that the users can upload their data in the cloud. But where the cloud storage is to be made secured.

However, the cloud is untrusted by users since the Cloud Service Providers are likely to be outside of the cloud users trusted domain. The cloud server will not maliciously delete or modify user data due to the protection of data auditing schemes but also will try to learn the content of the stored data and the identities of cloud users.

5.2 Verification Phase

At the time of verification process the verifier uses the proof of data integrity protocol and metadata for checking the integrity of the data. Whether the data is tampered or deleted. This phase is not prevent from the modifying the data for every file which is uploaded in the cloud. The verification process is completed by Third Party Authenticator (TPA).

5.3 Group Manager Module

Group Manager takes care of system parameter generation, User registration and User Revocation in the cloud. The group manager work as admin of a group so it knows the logs of each and every process in the cloud. For new user registration and revocation process the group manager is responsible.

5.4 Group Member Module

Group members is set of users which are registered within the same group so the can store their private data and also can share this stored data with other member in the same group. The group members are dynamically changed due to user revocation process. The existing group member can upload and view the files as well as modify the file within the same group.

5.5 File Security Module

This module can encrypt the data file with the group. The file stored in the cloud can deleted by group manager either data owner.(member who uploded the file on the cloud)

5.6 Ring Signature Module

A ring signature scheme allows any member of the group to sign messages while keeping the identity secret from verifiers. Besides, the designated rings manager cannot disclose the identity of the signature's thus maintaining the confidentiality of the user.

5.7 User Revocation Module

User revocation is performed by the group manager via available revocation list (RL), based on which group members can encrypt their data files and ensure the confidentiality against the revoked users.

6. Conclusion

In this paper to facilitate the client in getting a proof of integrity of the data which their wishes to store in the cloud storage servers with bare minimum expenses and efforts. This scheme is develop for reduce the computational and storage overhead of the client as well as to minimize the computational overhead of the cloud storage server also minimized the size of the proof of data integrity so as to reduce the network bandwidth consumption. As well as designed a secure data sharing scheme for dynamic groups in an untrusted cloud. A user is able to share data with others in the group without revealing identity privacy to the cloud. Additionally, this supports efficient user revocation and new user joining. More specially, efficient user revocation can be achieved through a public revocation list without updating the private keys of the remaining users, and new users can directly decrypt files stored in the cloud before their participation. Additionally, the storage overhead and the encryption computation cost are continuous. Extensive analyses show that proposed scheme satisfies the desired security requirements and guarantees efficiency as well.

References

[1] Privacy Protection Techniques in Cloud Computing International Journal of Application or Innovation in Engineering & Management (IJAIEM) Volume 2, Issue 12, December 2013 ISSN 2319 - 4847

[2] B. Wang, B. Li, and H. Li, "Oruta: Privacy-Preserving Public Auditing for Shared Data in the Cloud," Proc. IEEE Fifth Int'l Conf. Cloud Computing, pp. 295-302, 2012.

[3] 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.

[4] K. Ren, C. Wang, and Q. Wang, "Security Challenges for the Public Cloud," IEEE Internet Computing, vol. 16, no. 1, pp. 69-73, 2012.

[5] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving Secure, Scalable, and Fine-Grained Data Access Control in Cloud Computing," Proc. IEEE INFOCOM, pp. 534-542, 2010

[6] 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.

[7] B. Wang, M. Li, S.S. Chow, and H. Li, "Computing Encrypted Cloud Data Efficiently under Multiple Keys," Proc. IEEE Conf. Comm. and Network Security (CNS '13), pp. 90-99, 2013.

[8] R. Rivest, A. Shamir, and L. Adleman, "A Method for Obtaining Digital Signatures and Public Key Cryptosystems," Comm. ACM, vol. 21, no. 2, pp. 120-126, 1978.

[9] The MD5 Message-Digest Algorithm (RFC1321). https://tools. ietf.org/html/rfc1321, 2014.

[10] 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.

[11] H. Shacham and B. Waters, "Compact Proofs of Retrievability," Proc. 14th Int'l Conf. Theory and Application of Cryptology and Information Security: Advances in Cryptology (ASIACRYPT '08), pp. 90107, 2008.

[12] 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.

[13] 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.

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

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

[16] 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.

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

[18] B. Wang, B. Li, and H. Li, "Certificateless Public Auditing for Data Integrity in the Cloud," Proc. IEEE Conf. Comm. and Network Security (CNS'13), pp. 276-284, 2013.

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

[20] Jyothi Kannan, Leda Kamal, "Enhanced Technique for Privacy Preserving Public Auditing for Shared Data in Cloud", international Journal of Engineering Research & Technology (IJERT) ,ISSN: 2278-0181, IJERTV4IS030384, www.ijert.org Vol. 4 Issue 03, March-2015