Survey on Secure Data Deduplication in Cloud Environment

S.Vinitha[1], S. Prabhu[2]

[1]PG Scholar, Department of CSE, Nandha Engineering College (Autonomous), Erode, Tamil Nadu, India

[2]Professor, Department of CSE, Nandha Engineering College (Autonomous), Erode, Tamil Nadu, India.

Email: svinithadharshini@gmail.com[1], prabhu.s@nandhaengg.org[2]

Received date: 28th November, 2017, Revised Date: 15th December, 2017, Accepted Date: 24th December, 2017

Abstract - Cloud computing is an information technology (IT, a model for enabling ubiquitous access to shared pools of configurable resources. Data de duplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. Although convergent encryption has been extensively adopted for secure de duplication, a critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. This paper provides the comprehensive study of cloud security issues, various algorithms and encryption and decryption techniques used to ensure the secure data deduplication in cloud environment.

Keywords - Cloud computing, Storage, Techniques, Issues, encryption and decryption.

I. INTRODUCTION

Cloud computing has gained interest in various kinds of domains like business, education, research, market, publications etc. It is a technology which has shifted the cost of maintaining large servers which usually are under apply to third party vendors. Due to this shift many small and medium level organizations deploy their application with just the usage cost (pay-as-you-go). Cloud computing offers three standard models Infrastructure as a Service, Platform as a Service, and Software as a Service. The cloud deployment models are Private, Public and Hybrid.

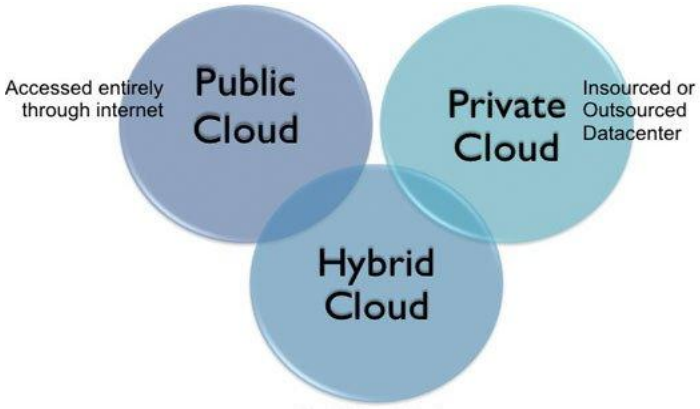

Fig. 1 Cloud Deployment Models

A. Types of cloud computing

- Infrastructure as a Service means you are selling access to raw computing hardware over the Net, such as servers or storage. Since you buy what you need and pay, this is often referred to as utility cloud computing. Ordinary web hosting is a simple example of pay a monthly subscription or a per megabyte fees to have a hosting company server up files for your website from other servers.
- Software as a Service means you have use a complete application running on someone other system. Web-based servers are email and Google Documents are perhaps the best-known examples. Other software is another well-known provider offering a variety of office applications online.
- Platform as a Service means you develop applications using Web-based tools so they run on system software and system hardware provided by another company. So, for example you might develop your own new website but have the whole thing, including the shopping cart, checkout and payment mechanism running on a merchant's server. Force.com and the Google App Engine are examples of Platform as a service.

B. Cloud Service

- A cloud service is a remotely accessible environment not all resources residing within a cloud can be made available for remote access. For example a database and physical server deployed within a cloud service may be only accessible by other resources that are within the same cloud.
- A cloud service is any resource that is made remotely accessible via a cloud. Unlike other fields that fall under the service technology umbrella such as service-oriented architecture the term service within the context of cloud computing especially broad.

C. Network Cloud Services

- What you may not have thought about is that every one of these consumer application cloud services uses network cloud services. In fact, the word "cloud" comes from the fact that many years ago those of us who built and sold client server applications, software and hardware used to draw a picture with the PC connected to a network and the network connected to a server.
- Since none of us actually understood how the network worked, we drew a cloud and labeled it "network" and left it at that. In those days companies built their own networks, but today consumers and businesses use network cloud services delivered by companies like AT&T, Verizon and Sprint.

D. Application Cloud Services

- So we have focused on consumer application cloud services, but for the past ten years the fastest-growing business applications have all been delivered as cloud services. Since 1999, fifteen companies that deliver business application cloud services have become public companies. These fifteen include Concur (1999), Webex (2000), Kintera (2003), Salesforce.com (2004), RightNow Technologies (2004), WebSideStory (2004), Kenexa (2005), Taleo (2005), DealerTrack (2005), Vocus (2005), Omniture (2006), Constant Contact (2007), SuccessFactors (2007), NetSuite (2007), and OpenTable (2009).

II. MATERIALS AND METHODS

The following literature survey shows the various techniques and algorithms which have been proposed to heighten the data security in cloud.

A. Secure Anti-Collusion Data Sharing Scheme for Dynamic Groups in the Cloud [1]

We design a secure anti-collusion data sharing scheme for dynamic groups in the cloud. In our scheme, the users can securely obtain their private keys from group manager Certificate Authorities and secure communication channels. Also, our scheme is able to support dynamic groups efficiently when a new user joins in the group or a user is revoked from the group the private keys of the other users do not need to be recomputed and updated. Moreover, our scheme can achieve secure user revocation the revoked users can not be able to get the original data files once they are revoked even if they conspire with the cloud.

B. Privacy Preserving Public Auditing for Data Storage Security [2]

The proposed system specifies that user can access the data on a cloud as if the local one without worrying about the integrity of the data. Hence, TPA is used to check the integrity of data. It supports privacy preserving public auditing. It checks the integrity of the data, storage correctness. It also supports data dynamics & batch auditing. The major benefits of storing data on a cloud is the relief of burden for storage management, universal data access with location independent & avoidance of capital expenditure on hardware, software & personal maintenance.

C. Data Security, Privacy, Availability and Integrity in Cloud Computing: Issues and Current Solutions [3]

The cost of this technology is more attractive when it is compared to building the infrastructure. However, there are many security issues coming with this technology as happens when every technology matures. Those issues include issues related to the previous issues of the internet, network issues, application issues, and storage issues. Storing data in a remote server leads to some security issues. Those issues are related to confidentiality of data from unauthorized people in remote sites, integrity of stored data in remote servers and the availability of the data when it is needed. Also, sharing data in cloud when the cloud service provider is mistrusted is an issue.

D. Secure Auditing and Deduplicating Data in Cloud [4]

Achieving both data integrity and deduplication in cloud, we propose SecCloud and SecCloud+. SecCloud introduces an auditing entity with maintenance of a MapReduce cloud, which helps clients generate data tags before uploading as well as audit the integrity of data having been stored in cloud. In addition, SecCoud enables secure deduplication through introducing a Proof of Ownership protocol and preventing the leakage of side channel information in data deduplication. Compared with previous work, the computation by user in SecCloud is greatly reduced during the file uploading and auditing phases. SecCloud+ is an advanced construction motivated by the fact that customers always want to encrypt their data before uploading, and allows for integrity auditing and secure deduplication directly on encrypted data.

E. Secure Data Deduplication with Dynamic Ownership Management in Cloud Storage [5]

Features a reencryption technique that enables dynamic updates upon any ownership changes in the cloud storage. Whenever an ownership change occurs in the ownership group of outsourced data, the data are reencrypted with an immediately updated ownership group key, which is securely delivered only to the valid owners. Thus, the proposed scheme enhances data privacy and confidentiality in cloud storage against any users who do not have valid ownership of the data, as well as against an honest-but-curious cloud server. Tag consistency is also guaranteed, while the scheme allows full advantage to be taken of efficient data deduplication over encrypted data. In terms of the communication cost, the proposed scheme is more efficient than the previous schemes, while in terms of the computation cost, taking additional 0:1 □ 0:2 ms compared to the RCE scheme, which is negligible in practice.

F. Privacy-Preserving Public Auditing for Regenerating-Code- Based Cloud Storage [6]

Considering that the data owner cannot always stay online in practise, in order to keep the storage available and verifiable after a malicious corruption, we introduce a semi trusted proxy into the system model and provide a privilege for the proxy to handle the reparation of the coded blocks and authenticators. To better appropriate for the regenerating-codescenario, we mapping our authenticator based on the BLS signature. This authenticator can be efficiently generated by the data owner simultaneously with the encoding procedure. Extensive analysis shows that our scheme is provable secure, and the performance evaluation shows that our scheme is highly efficient and can be feasibly integrated into a regenerating-code-based cloud storage system.

G. Privacy-Preserving Public Auditing Protocol for Low Performance End Devices in Cloud [7]

Privacy-preserving public auditing protocols for secure storage in cloud environment. Our protocols are based on online/offline signatures, by which a user only needs to perform lightweight computing when a data file to be outsourced is given. Further, our protocols also support batch auditing and data dynamics. Simulation shows that our protocol is much more efficient than a recent privacypreserving public auditing protocol. Thus, we believe that our protocols are practical for those end devices with low computation capabilities.

H. Mutual Verifiable Provable Data Auditing in Public Cloud Storage [8]

We define MV-PDP system model and security model. And then we utilize of Diffie-Hellman shared key to construct the homomorphic authenticator. In MV-PDP system, the data blocks signed by a client can be verified by a private verifier, while the data blocks signed by a verifier can also be check by a client. And the same data blocks are easy to be signed and checked by a client and the verifier in turn. Furthermore, in MV-PDP the verifier is stateless and independent of CSS. We’ d like to emphasize that ECC-based homomorphic authenticator is used to design our scheme, which result in low calculation and Communications due to the fact that bilinear operation is not required.

I. Privacy-Preserving Public Auditing for Regenerating-Code-Based Cloud Storage [9]

Considering that the information proprietor can't generally stay online practically speaking, with a specific end goal to keep the capacity accessible and variable after a malevolent debasement, I bring a semi-trusted intermediary into the framework demonstrate and give a benefit to the intermediary to handle the reparation of the coded pieces and authenticators. Broad examination demonstrates that these proposed plan is provable secure, and the execution assessment will demonstrate that propose plan is exceedingly effective and can be possibly incorporated into a recovering code-based distributed storage framework.

J. Public Auditing for Shared Data with Efficient User Revocation in the Cloud [10]

New public auditing mechanism for shared data with efficient user revocation in the cloud. When a user in the group is revoked, we allow the semi-trusted cloud to re-sign blocks that were signed by the revoked user with proxy re-signatures. Experimental results show that the cloud can improve the efficiency of user revocation, and existing users in the group can save a significant amount of computation and communication resources during user revocation.

III. RESULTS AND DISCUSSION

The following table summarizes different algorithms are working on storage parameters at some cases. Each algorithm focuses on improving different parts of cloud environment .The differences are shown in Table I

TABLE I: DIFFERENT TECHNIQUES & IMPACTS

Sl.	Techniques and Algorithms	Impacts
1	Watermarking technique for data Storage	Our scheme can achieve fine efficiency.
2	RBAC scheme.	Revoked users can not be able to get the original data files once they are revoked even if they conspire with the untrusted cloud.
3	MAC algorithm Message Authentication Code	More secure cloud storage and lead to more acceptance from the people.

4	SecCloud and SecCloud+.	Encrypt their data before uploading, and enables integrity auditing and secure deduplication on encrypted data.
5	Demonstrate the feasibility of the proposed scheme	Scheme to provide a mutual nonrepudiation guarantee in their service-level agreements.
6	OEAP Algorithm	Authenticator can be efficiently generated by the data owner simultaneously with the encoding procedure.
7	Online/offline signatures	User only needs to perform lightweight computing when a data file to be outsourced is given.
8	Diffie-Hellman shared key	Low calculation and Communications due to the fact that bilinear operation is not required.
9	Cryptography framework	Exceedingly effective and can be possibly incorporated into a recovering code-based distributed storage framework.
10	Data with efficient user revocation.	Group can save a significant amount of computation and communication resources during user revocation.

IV. CONCLUSION

The utilization of cloud computing paradigm is continuously growing. Cloud storage is the one of the most important task in the project also considers the revocation of users in the given group. If the original (first) user of the group intimates the server with a user's (B) revocation, then the server rejects the proof of ownership submitted by that user (B). Likewise, session based deduplication is considered. Here if the user provides the session duration i.e, front date and to date, then only with the data range, proof of ownership can be allowed in server on those dates. This increases the security if the outsourced data need to be safely accessed on the given duration.

ACKNOWLEDGMENT

I am thankful for the timely and consistent cooperation given by my guide S. Prabhu for preparing this survey. I hope this survey will help to understand various kinds of cloud storage threats and techniques available with the aspect of secure cloud platform.

REFERENCES

[1] Tejashree Paigude and T. A. Chavan, A Secure Anti Collusion Data Sharing Scheme for Dynamic Groups in the Cloud, IEEE Transactions on Parallel and Distributed Systems, Vol. 27, No. 1, pp. 40-50, 2016.
[2] Sultan Aldossary and William Allen, Data Security, Privacy, Availability and Integrity in Cloud Computing: Issues and Current Solutions, International Journal of Advanced Computer Science and Applications, Vol. 7, No. 4, pp. 485-498, 2016.
[3] Jingwei Li, Jin Li, Dongqing Xie and Zhang Cai, Secure Auditing and Deduplicating Data in Cloud, IEEE Translations on Computers, Vol. 65, No. 8, pp. 2386-2396, 2016.
[4] Gwan-Hwan Hwang and Hung-Fu Chen, Efficient Real-time Auditing and Proof of Violation for Cloud Storage Systems, IEEE 9th International Conference on Cloud Computing, DOI: 10.1109/CLOUD.2016.0027, 2016.
[5] A. P. Mohana Priyaa and A. Gokilavani, Privacy-Preserving Public Auditing for Regenerating-Code-Based Cloud Storage, IEEE Transactions on Information Forensics and Security, Vol. 10, No. 7, pp. 1513-1528, 2015
[6] Jiangtao Li, Lei Zhang, Joseph K. Liu, Haifeng Qian and Zheming Dong, Privacy-Preserving Public Auditing Protocol for Low Performance End Devices in Cloud, IEEE Transactions on Information Forensics and Security, Vol. 11, No. 11, pp. 2572-2583, 2016.
[7] Jin Wang, Mutual Verifiable Provable Data Auditing in Public Cloud Storage, Journal of Internet Technology, Vol. 16, No. 2, pp. 317-323, 2015.
[8] Boyang Wang, Baochun Li and Hui Li, Public Auditing for Shared Data with Efficient User Revocation in the Cloud, IEEE Transactions on Services Computing, Vol. 8, No. 1, pp. 92-106, 2015.