Leveraging De-duplication with Secure Auditing in Cloud Data

Vidya Patil[1], Prof.Aruna Verma[2]

[1,2]Department of computer engineering, Dhole patil college of Engineering, Pune

Abstract- **secure data deduplication can significantly reduce the storage overheads in cloud storage services. specially aim to achieving data integrity and deduplication in cloud.in existing data deduplication schemes are generally designed to either resist brute force attacks or ensure the efficiency and data availability but not both conditions.in this seminar propose two secure system SecCloud and SecCloud+ seccloud maintain mapreduce cloud which help clients generates data tages before uploading as well as audit the integrity of data have been stored in cloud.seccloud reduced the file uploading and auditing phases. seccloud+ is design for encrypted the customers data before uploading and enables integrity auditing. this are achieve both the privacy preserving and data availability.**

Index Terms- Secure Data De-duplication, Encrypted Data, Data Availability, accountability, Cloud Computing.

1. INTRODUCTION

Cloud storage are offers a new way of information technology sevices by rearranging various resources e,g. storage ,computing and providing them to users based on their demands. clous computing provide a big resourses pool by linking network resoures together. User uploads the data to the cloud, require this data securely store so use cryptography algorithm for security purpose. to Cloud storage is a model of internet enterprise storage where data is stored in virtualized pools of storage which is hosted by third-party. Cloud storage provides offers for customer which generated more benefit for cloud companies, like popularity, more user. Even though now days cloud storage system has been smart option for work. And also it is affordable, but it has certain limitation. The main problem of client data management and maintenance which is able to Relief by cloud server storage system of cloud is different from another storage System. The first problem is integrity auditing, i.e when we uploaded data it upload various manner like packets tokens which is less secure because if any packet loss while transmitting it's occur problem for client. As well as its to easy for a professional Attacker to attack. So its most important that maintain the integrity of data on storage system. The data is transferred via internet and stored in uncertain domain not the under control of client. The uncontrolled cloud server may passively hide the any problem related data for their reputation. It is more important that cloud server might even actively and deliberately discard rarely accessed data files belonging to an ordinary file. The second problem is secure deduplication. In cloud storage among these remote stored files, most of them are already on storage. According to recent survey by EMC, 70% of files are duplicated copies. Because its helps to cloud servers paid more for space from client. Thats the one of the reason why many cloud server are store duplicate copies of data. And Its more risky to available duplictate copies of data in storage.

Stored data is various manners like confidential password, banking detail, personal information, it is open invitation for attacker. In cloud server, server stores every single file link with the who ask for the file. Cloud server needs to verify whether the user actually owns the file before creating a link for user. In de-duplicate data, when a user wants to upload a data file that already exists in the cloud storage, the cloud server executes a checking algorithm to see whether or not this user actually possesses the whole file i.e. it checks the file attribute. If the user passes the checking, he/she can directly use the file existed on the server without uploading it again. To overcome such problems cloud server uses proofs-of-ownership protocol, which let a client efficiently prove to a server that the client holds a file, rather than short information about it. In this a file have different ownership which introduce rigorous security definition. In cloud computing technology we can access any file anywhere ,anytime.

2. LITERATURE SURVEY

The author S.Keelveedhi ,M.Bellare propose dupless concept that combines a MLE Scheme and CEtype with the ability to obtain message deriverd keys with the help of a key server shared to the group of clients.the client interact with the KS bt a protocol for oblivious PRFs,that the Ks can cryptographically mix in secret material to the pre message keys while learning nothing about files stored by clients[4]

The author H.wang, proposed the ppdp concept.it provide the security to the clients. author Gives the security model and system model and design the efficient pairing based ppdp protocol.this ppdp protocol is provably secure and efficient by security analysis and performance analysis[5]

The author J. Li, X. Tan, X. Chen, and D. Wong proposed a new cloud storage architecture with two independent cloud servers, that is, the cloud storage server and the cloud audit server. The cloud audit server allows cloud users, to pre-process the data before uploading to the cloud storage server and verify data integrity. The cloud audit server eliminates the involvement of user in the auditing and in the pre-processing phases[6]

The author M. Bellare, S. Keelveedhi, and T. Ristenpart, formalize a new cryptographic primitive, Message-Locked Encryption where the key under which encryption and decryption are performed is itself derived from the message. MLE provides a way to achieve secure de-duplication a goal currently targeted by numerous cloud-storage providers[7]

The author M. Abadi, D. Boneh, I. Mironov, A. Raghunathan, and G. Segev propose, achieve the goal via a combination of a cut-and-choose technique and NIZKs. The resulting scheme is secure against a fully adaptive adversary. The second construction assumes a predetermined bound on the complexity of distributions specified by the adversary. It fits the original framework of deterministic MLE while satisfying a stronger security notion[8]

The author M. Azraoui, K. Elkhiyaoui, R. Molva, and M. Onen propose, Stealth Guard, an efficient and provably secure proof of retriev abillity (POR) scheme. Stealth Guard makes use of a privacy preserving word search (WS) algorithm to search, as part of a POR query, for randomly valued blocks called watchdogs that are inserted in the file before outsourcing[9]

3. SYSTEM ARCHITECTURE

The system model comprises a key distribution center, cloud service provider, clients from different domains and the corresponding local manager, denoted by lma and lmb. Cloud service does not modify the stored message due to reputation. The lma and lmb is curious about the data uploaded by staff however to protect the information asset lma or lmb does not actively attempt to compromise the privacy of client. in drop concept divide a file into fragments and replicate a fragmented data over the cloud node. each of the node store only a single fragment of a particular data file that ensures even in case of a successful attack, no meaningful information is revealed to the attacker.

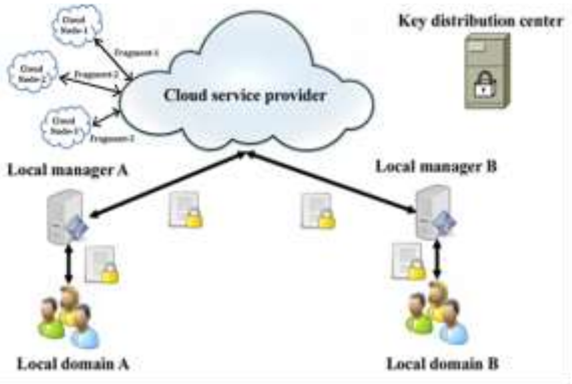


Figure1. System architecture

3.1 Key Distribution Center

The trusted kdc is worked in distribution and management of private keys for the system.

3.2 Clients

Clients upload and save their data with the csp.in order to protect their data privacy and help the csp to complete data deduplication over encrypted data. they encrypt the data and generate the corresponding tags.

3.3 LMA and LMB

Each domain maintain a local manager example lma and lmb which is responsible for intra de duplication and forwarding message from clients in domain a or b.

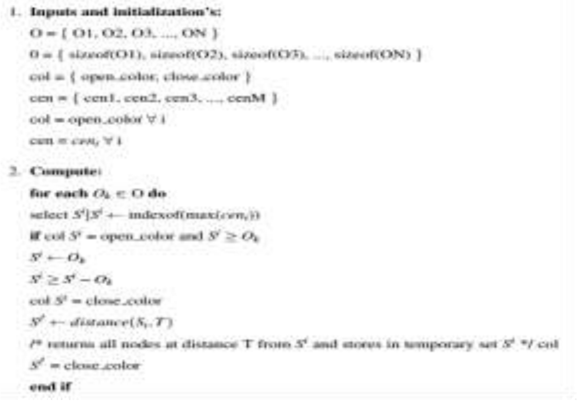
3.4 Cloud Servers

Cloud Servers virtualizes the resources according to the requirements of clients and expose them as storage pools. This system uses Division and Replication of Data in the Cloud for Optimal Performance and Security (DROPS) that fragments user files into pieces and replicates them at strategic node locations within the cloud. Typically, the cloud clients may buy or lease storage capacity from cloud servers, and store their individual data in these bought or rented spaces for future utilization.

4. ALGORITHM

4.1 Fragment Placement Algorithm

Divide the files into fragments and replicated the fragmented data over the cloud nodes. This algorithm represents the fragment placement methodology. To deal with the security aspects of placing fragments, this system use the concept of T-coloring that was originally used or the channel assignment problem. It generates a non-negative random number and builds the set T starting from zero to the generated random number. The set T is used to restrict the node selection to those nodes that are at hop-distances not belonging to T. For the said purpose, it assigns colors to the nodes, such that, initially, all of the nodes are given the open color. Once a fragment is placed on the node, all of the nodes within the neighborhood at a distance belonging to T are assigned close color. In the aforesaid process, this loses some of the central nodes that may increase the retrieval time but it achieves a higher security level. If somehow the intruder compromises a node and obtains a fragment, then the location of the other fragments cannot be determined. The attacker can only keep on guessing the location of the other fragments.



```
1. Inputs and initialization's:
   O = { O1, O2, O3, ...., ON }
   0 = { sizeof(O1), sizeof(O2), sizeof(O3), ...., sizeof(ON) }
   col = { open_color, close_color }
   cen = { cen1, cen2, cen3, ...., cenM }
   col = open_color ∀ i
   cen = $cen_i$ ∀ i

2. Compute:
   for each $O_k \in O$ do
   select $S^i | S^i \leftarrow$ indexof(max($cen_i$))
   if col $S^i$ = open_color and $S^i \geq O_k$
   $S^i \leftarrow O_k$
   $S^i \geq S^i - O_k$
   col $S^i$ = close_color
   $S^{i'} \leftarrow distance(S_i, T)$
   /* returns all nodes at distance T from $S^i$ and stores in temporary set $S^{i'}$ */ col
   $S^{i'}$ = close_color
   end if
```

5. CONCLUSION AND FUTURE WORK

This work present achieving efficient and privacy preserving data deduplication in cloud storage. SecCloud and SecCloud this two secure system are use.ths two system are provide the security to the data for the users.Seccloud introduce the auditing entity and maintain the mapreduce cloud. which help clients generstes data tags before uploading and audit the integrity of data having been stored in cloud.SecCloud+ design for the customer wants to encrypt their data before uploading and allows the integrity auditing and secure deduplication directly on encrypted data. .Drop concept use for reduce storage and bandwidth.this are divide the file into fragment and replicate the fragmented data over the cloud nodes. This are achieving both data availability and privacy preserving.

REFERENCES

[1] Y. Zhu, H. Hu, G.-J. Ahn, and M. Yu, “Cooperative provable data possession for integrity verification in multicloud storage”, IEEE Transactions on Parallel and Distributed Systems, 2012.

[2] J. Xu and E.-C. Chang, “Towards efficient proofs of retrievability”, in Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security, 2012.

[3] J. Yuan and S. Yu, “Secure and constant cost public cloud storage auditing with deduplication”, in IEEE Conference on Communications and Network Security (CNS), 2013.

[4] S. Keelveedhi, M. Bellare, and T. Ristenpart, “Dupless: Server aided encryption for deduplicated storage”, in Proceedings of the 22Nd USENIX Conference on Security, ser. SEC13. Washington, D.C.: USENIX Association, 2013.

[5] H. Wang, “Proxy provable data possession in public clouds”, IEEE Transactions on Services Computing, vol. 6, no. 4, pp. 551559, 2013.

[6] J. Li, X. Tan, X. Chen, and D. Wong, “An efficient proof of retrievability with public auditing in cloud computing”, in 5th International Conference on Intelligent

Networking and Collaborative Systems (INCoS), 2013.

[7] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication", in Advances in Cryptology EUROCRYPT 2013.

[8] M. Abadi, D. Boneh, I. Mironov, A. Raghunathan, and G. Segev, "Message locked encryption for lock-dependent messages", in Advances in Cryptology CRYPTO 2013.

[9] M. Azraoui, K. Elkhiyaoui, R. Molva, and M. Onen, "Stealthguard: Proofs of retrievability with hidden watchdogs", in Computer Security ESORICS 2014.

[10] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management", IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 6, pp. 16151625, June 2014.