International Journal of Trend in Scientific Research and Development (IJTSRD)
International Open Access Journal
ISSN No: 2456 - 6470 | www.ijtsrd.com | Volume - 2 | Issue – 2

User Data Integrating with Anti-Collusion Scheme and SVC in Cloud Groups

Meenambigai. D
Department of CSE, Mailam College of Engineering, Villupuram, Tamil Nadu, India

Mathavan. V
Associate Professor, CSE, Mailam College of Engineering, Villupuram, Tamil Nadu, India

ABSTRACT

In cloud computing, user can share data among group members with the characters of less maintenance and little management cost. Sharing data must have security guarantees, if they are out sourced. Sharing data while providing privacy preserving is still a challenging problem, when change of the membership. It might cause to the collusion attack for an unsecured cloud. For existing technique, security of key distribution is based on the secure communication channel, however, to have such channel is a strong assumption and is difficult for practice. We propose a secure data sharing scheme for dynamic users. Key distribution done without any secure communication channels and the user can get the individual key from group manager. Data deduplication is one of the techniques which used to solve the repetition of data. Our proposed system prevents the replication of files and media file like images, videos. The deduplication techniques are generally used in the cloud server for reducing the space of the server. To prevent the unauthorized use of data accessing and create duplicate data on cloud the encryption technique to encrypt the data before stored on cloud server. CloudMe is proposed for cloud storage. All files of data owners are encrypted using AES algorithm and stored in real cloud. Thus we present a secure system architecture design as our initial effort towards this direction, which bridges together the advancements of video coding techniques and secure deduplication. Our design enables the cloud with the crucial deduplication functionality to completely eliminate the extra storage and bandwidth cost.

Keywords: *Cloud Computing, Privacy Preserving, Anti-Collusion, Deduplication, Key distribution*

I. INTRODUCTION

Cloud Computing is an innovative technology that is revolutionizing the way we do computing. The key concept of cloud computing is that you don't buy the hardware, or even the software, you need anymore, rather you rent some computational power, storage, databases, and any other resource you need by a provider according to a pay-as-you-go model, making your investment smaller and oriented to operations rather than to assets acquisition.

Cloud computing can be defined as a model for enabling ubiquitous, convenient and on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort from the user side and minimal service provider interaction. Cloud computing is considered the evolution of a variety of technologies that have come together to change an organizations' approach for building their IT infrastructure. Actually, there is nothing new in any of the technologies that are used in the cloud computing where most of these technologies have been known for ages. It is all about making them all accessible to the masses under the name of cloud computing. Cloud is not simply the latest term for the Internet, though the Internet is a necessary foundation for the cloud, the cloud is something more than the Internet. The cloud is where you go to use technology when you need it, for as long as you need it. You do

not install anything on your desktop, and you do not pay for the technology when you are not using it. The cloud can be both software and infrastructure. It can be an application you access through the Web or a server like Gmail and it can be also an IT infrastructure that can be used as per user's request. Whether a service is software or hardware, the following is a simple test to determine whether that service is a cloud service.

In the simplest terms, cloud computing means storing and accessing data and programs over the Internet instead of your computer's hard drive. The cloud is just a metaphor for the Internet. It goes back to the days of flowcharts and presentations that would represent the gigantic server-farm infrastructure of the Internet as nothing but a puffy, white cumulus cloud, accepting connections and doling out information as it floats.

Hybrid services like Box, Dropbox, and SugarSync all say they work in the cloud because they store a synced version of your files online, but they also sync those files with local storage. Synchronization is a cornerstone of the cloud computing experience, even if you do access the file locally.

The size of Cloud storage is expanding at a dramatic speed. It is estimated that by 2015 the data stored in the Cloud will reach 0.8 ZB while even more data is "touched" by the Cloud within the data lifecycle. Meanwhile, with the development of the Cloud computing paradigm, Cloudbased applications have put forward a higher demand for Cloud storage. While the requirement of data reliability should be met in the first place, data in the Cloud needs to be stored in a highly cost-effective manner. In this paper, our research focuses on minimizing the Cloud storage consumption by minimizing data replication while meeting the data reliability requirement.

Data deduplication is used in our scheme which is one of the techniques which used to solve the repetition of data. The deduplication techniques are generally used in the cloud server for reducing the space of the server. To prevent the unauthorized use of data accessing and create duplicate data on cloud the encryption technique to encrypt the data before storedon cloud server.

In our scheme we propose anti–collusion scheme in cloud groups and integrated data deduplication technique. The following are the major contribution of our scheme:

1) In our scheme we provide, secure protocol for anti -collusion attack. It does not allow revoked user to access files in the cloud. There is no need for recomputing or updating private keys of other users.
2) By integrating algorithms / techniques we can implement deduplication concepts and reduce the storage cost in a cloud for the data owners.
3) Faster recovery and processing of data. This would significantly decrease the processing time of load balancer.
4) Effective and Efficient usage of cloud Storage Space. As data deduplication technique is used in this scheme the cloud storage is used effectively as there will not be any file of same content. Deduplication technique checks the content of each file while user uploads new files into the cloud.
5) Our scheme includes privacy preserving environment, only legal user can access through groups, any irrelevant entity cannot recognize the exchanged data and communication state even it intercepts the exchanged messages via an open channel, any adversary cannot correlate two communication sessions to derive the prior interrogations according to the currently captured messages.

II. LITERATURE SURVEY

A.Juels [1] in this paper,we define and explore proofs of retrievability (PORs). A POR scheme enables an archive or back-up service (prover) to produce a concise proof that a user (verifier) can retrieve a target file F, that is, that the archive retains and reliably transmits file data sufficient for the user to recover Fin its entirety.A POR may be viewed as a kind of cryptographic proof of knowledge (POK), but one specially designed to handle a large file (or bitstring) F. We explore POR protocols here in which the communication costs, number of memory accesses for the prover, and storage requirements of the user (verifier) are small parameters essentially independent of the length of F. In addition to proposing new, practical POR constructions, we explore implementation considerations and optimizations that bear on previously explored, related schemes. In a POR, unlike a POK, neither the prover nor the verifier need actually have knowledge of F. PORs give rise to

a new and unusual security definition whose formulation is another contribution of our work. We view PORs as an important tool for semi-trusted online archives. Existing cryptographic techniques help users ensure the privacy and integrity of files they retrieve. It is also natural, however, for users to want to verify that archives do not delete or modify files prior to retrieval. The goal of a POR is to accomplish these checks without users having to download the files themselves. A POR can also provide quality-of-service guarantees, i.e., show that a file is retrievable within a certain time bound.

G. Ateniese [2] Storage outsourcing is a rising trend which prompts a number of interesting security issues, many of which have been extensively investigated in the past. However, Provable Data Possession (PDP) is a topic that has only recently appeared in the research literature. The main issue is how to frequently, efficiently and securely verify that a storage server is faithfully storing its client's (potentially very large) outsourced data. The storage server is assumed to be untrusted in terms of both security and reliability. (In other words, it might maliciously or accidentally erase hosted data; it might also relegate it to slow or off-line storage.) The problem is exacerbated by the client being a small computing device with limited resources. Prior work has addressed this problem using either public key cryptography or requiring the client to outsource its data in encrypted form. In this paper, we construct a highly efficient and provably secure PDP technique based entirely on symmetric key cryptography, while not requiring any bulk encryption. Also, in contrast with its predecessors, our PDP technique allows outsourcing of dynamic data, i.e, it efficiently supports operations, such as block modification, deletion and append.

F. Chen [3] Empowered by today's rich tools for media generation and distribution, and the convenient Internet access , streaming crowdsourced multimedia content (crowdsourced streaming, in brief) generalizes the single-source streaming paradigm by including massive contributors for a video/data channel. It calls a joint optimization along the path from crowd sourcers, through streaming servers, to the end-users to minimize the overall latency. The dynamics of the video sources, together with the globalized request demands and the high computation demand from each sourcer, make crowdsourced live streaming challenging even with powerful support from modern cloud computing. In this paper, we present a generic framework that facilitates a cost-effective cloud service for crowdsourced live streaming. Through adaptively leasing, the cloud servers can be provisioned in a fine granularity to accommodate geo-distributed video crowdsourcers. We present an optimal solution to deal with service migration among cloud instances of diverse lease prices. It also addresses the location impact to the streaming quality. To understand the performance of the proposed strategies in the real world, we have built a prototype system running over the planetlab and the Amazon/Microsoft Cloud. Our extensive experiments demonstrate that the effectiveness of our solution in terms of deployment cost and streaming quality.

Y. Zheng [4] Multimedia contents, especially videos, are being exponentially generated today. Due to the limited local storage, people are willing to store the videos at the remote cloud media center for its low cost and scalable storage. However, videos may have to be encrypted before outsourcing for privacy concerns. For practical purposes, the cloud media center should also provide the deduplication functionality to eliminate the storage and bandwidth redundancy, and adaptively disseminate videos to heterogeneous networks and different devices to ensure the quality of service. In light of the observations, we present a secure architecture enabling the encrypted cloud media center. It builds on top of latest advancements on secure deduplication and video coding techniques, with fully functional system implementations on encrypted video deduplication and adaptive video dissemination services. Specifically, to support efficient adaptive dissemination, we utilize the scalable video coding (SVC) techniques and propose a tailored layer-level secure deduplication strategy to be compatible with the internal structure of SVC. Accordingly, we adopt a structure-compatible encryption mechanism and optimize the way how encrypted SVC videos are stored for fast retrieval and efficient dissemination. We thoroughly analyze the security strength of our system design with strong video protection. Furthermore, we give a prototype implementation with encrypted end to end deployment on Amazon cloud platform.

III. PROPOSED SYSTEM

The users can securely obtain their private keys from group manager. Users send request to group manager for access the wanted group, at that time our system

provide individual secure key to user without activation. Then group manager see the requests and activate the keys after confirm them. After user's private key gets activation, then only user can access the group. Our scheme have fine-grained access control, any user in the group can use the source in the cloud and revoked users cannot access the cloud again after they are revoked. In our proposed system the group manager performs the below tasks when an new user joins the group or a user has left the particular group, Update the whole user name list. Generate a secure key and encrypt the key without activation and send to the updated user list. Update the rights in the cloud server.

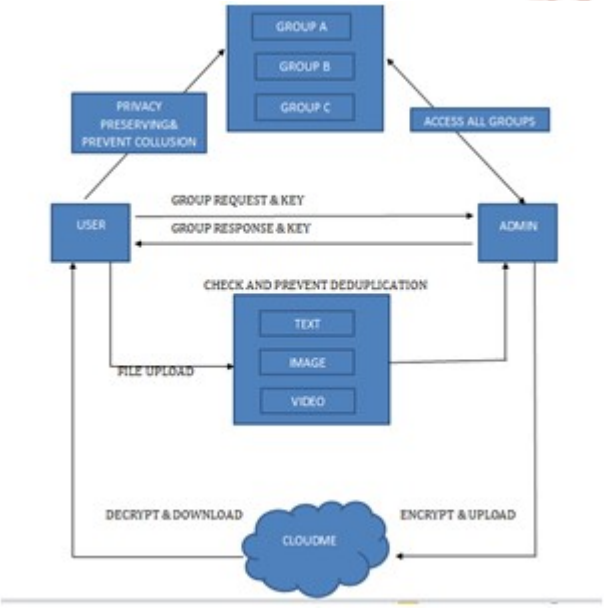

Fig.1 system architecture

PRIVACY PRESERVING

In this paper we address the issue of privacy preserving data mining. Specifically, we consider a scenario in which two parties owning confidential databases wish to run a data mining algorithm on the union of their databases, without revealing any unnecessary information. The privacy preserving data mining techniques are classified based on distortion, association rule, hide association rule, taxonomy, clustering, associative classification, outsourced data mining, distributed, and k-anonymity, where their notable advantages and disadvantages are emphasized.

KNN

In pattern recognition, the k-nearest neighbors' algorithm (KNN) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. KNN is a type of instance-based learning, or lazy learning. Where the function is only approximated locally and all computation is deferred until classification. The KNN algorithm is among the simplest of all machine learning algorithms.

DEDUPLICATION

Deduplication is a process to improve data quality by removing redundant or repetitive information from data in storage to improve storage utilization, simplify ETL, and optimize data transfers. Organizations often do not have visibility into the sources or causes of redundant data. Thus they have no way of knowing how much redundant data is costing them. For example, a retailer can waste a lot of money sending multiple copies of the same catalog or campaign to one prospective customer. By deduplicating the data ahead of time the company can prevent waste.

SVC

Scalable Video Coding (SVC) is the name for the Annex G extension of the H.264/MPEG-4 AVC video compression standard. SVC standardizes the encoding of a high-quality video bit stream that also contains one or more subset bit streams. A subset video bit stream is derived by dropping packets from the larger video to reduce the bandwidth required for the subset bit stream. The subset bitstream can represent a lower spatial resolution (smaller screen), lower temporal resolution (lower frame rate), or lower quality video signal. H.264/MPEG-4 AVC was developed jointly by ITU-T and ISO/IEC JTC 1. These two groups created the Joint Video Team (JVT) to develop the H.264/MPEG-4 AVC standard.

ADVANCED ENCRYPTION STANDARD (AES)

The Advanced Encryption Standard (AES), also known as Rijndael (its original name), is a specification for the encryption of electronic data established by the U.S. National Institute of Standards and Technology (NIST) in 2001. AES is a subset of the Rijndael cipher developed by two Belgian cryptographers, JoanDaemen and Vincen t Rijmen, who submitted a proposal to NIST during the AES selection process. Rijndael is a family of ciphers with different key and block sizes.

AES operates on a 4 × 4 column-major order matrix of bytes, termed the state, although some versions of Rijndael have a larger block size and have additional

columns in the state. Most AES calculations are done in a special finite field. AES consists of several rounds of several processing steps that include substitution, transposition and mixing of the input plaintext and transform it into the final output of ciphertext.

AES : Pseudocode

Cipher(byte in[16], byte out[16], key_array round_key[Nr+1])

```
begin
byte state[16];
state = in;
AddRoundKey(state, round_key[0]);
for i = 1 to Nr-1 stepsize 1 do
SubBytes(state);
ShiftRows(state);
MixColumns(state);
AddRoundKey(state, round_key[i]);
end for
SubBytes(state);
ShiftRows(state);
AddRoundKey(state, round_key[Nr]);
End
```

CLOUD STORAGE

Cloud storage is a model of data storage where the digital data is stored in logical pools, the physical storage spans multiple servers (and often locations), and the physical environment is typically owned and managed by a hosting company. These cloud storage providers are responsible for keeping the data available and accessible, and the physical environment protected and running. People and organizations buy or lease storage capacity from the providers to store user, organization, or application data.

The group user can upload the files in real cloud server named cloudMe. Duplication of files are checked and the files is been uploaded in the cloud server. To get a file, the user needs to send a request to the cloud server. The cloud server will also check the user's identity before issuing the corresponding file to the user. During file access the user key has to match by the group manager and the requested file can be downloaded by the group users.

IV. SYSTEM IMPLEMENTATION

The system output is mainly based on privacy preserving method. It will be evaluated using SVC algorithm. We propose a secure data sharing scheme for dynamic users. Key distribution done without any secure communication channels and the user can get the individual key from group manager. Data deduplication is one of the techniques which used to solve the repetition of data. By using this technique we prevents the replication of files and media file like images, videos. A secure protocol for anti-collusion attack are provided and Faster recovery and processing of data.

V. CONCULSION AND FUTURE WORK

Our proposed scheme provides a possible way to fight against immoral interference with the right of privacy. We hope more schemes can be created to protect cloud user privacy. Dynamic ownership management is an important and challenging issue in secure deduplication over encrypted data in cloud storage. Security analysis demonstrates that our schemes are secure in terms of insider and outsider attacks specified in the proposed security model. Also our proposed system effectively overcomes the collusion attack. Also prevention of storing a same file like text file, image and videos will optimize the user storage space and saves user storage space and cost.

VI. REFERENCES

1. R. Lu, X. Lin, X. Liang, and X. Shen, "Secure provenance: Theessential of bread and butter of data forensics in cloud computing," in Proc. ACM Symp. Inf., Comput. Commun. Security,2010, pp. 282–292.
2. B. Waters, "Ciphertext-policy attribute-based encryption: Anexpressive, efficient, and provably secure realization," in Proc. Int. Conf. Practice Theory Public Key Cryptography Conf. Public Key Cryptography,2008, pp. 53–70.
3. X. Liu, Y. Zhang, B. Wang, and J. Yang, "Mona: Secure multiowner data sharing for dynamic groups in the cloud," IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 6, pp. 1182–1191, Jun. 2013.
4. Z. Zhu, Z. Jiang, and R. Jiang, "The attack on mona: Secure multiowner data sharing for dynamic groups in the cloud," in Proc. Int. Conf. Inf. Sci. Cloud Comput., Dec. 7, 2013, pp. 185–189.

5. Zhongma Zhu and Rui Jiang, "A Secure Anti-Collusion Data Sharing Schemefor Dynamic Groups in the Cloud", IEEE Transactions On Parallel And Distributed Systems, Vol. 27, No. 1, January 2016
6. L. Zhou, V. Varadharajan, and M. Hitchens, "Achieving secure role-based access control on encrypted data in cloud storage," IEEE Trans. Inf. Forensics Security, vol. 8, no. 12, pp. 1947–1960, Dec. 2013.
7. D. T. Meyer and W. J. Bolosky, "A study of practical deduplication,"Proceedings of the 9th USENIX conference on File and storage technologies, ser. FAST'11. Berkeley, CA, USA: USENIX Association, pp. 1–6, 2011.
8. Moise´s G. de Carvalho, Alberto H.F. Laender, MarcosAndre´ Goncalves, and Altigran S. da Silva, "A Genetic Programming Approach to Record Deduplication," IEEE Transactions On Knowledge And Data Engineering, Vol. 24, No.3,pp.399 - 412, March 2012.
9. P.Shanmugavadivu,N.Baskar, "An Improving Genetic Programming Approach Based Deduplication Using KFINDMR",International Journal of Computer Trends and Technology, Vol.3, Issue5, pp.694-701, 2012.
10. W. Hu, T. Yang, J.N. Matthews, "The good, the bad and the ugly of consumer cloud storage," SIGOPS Oper. Syst. Rev. 44 pp.110–115,2010.
11. Amrita Upadhyay,Pratibha R Balihalli, ShashibhushanIvaturi andShrisha Rao, "Deduplication and Compression Techniques in Cloud Design", Proceedings of IEEE International Systems Conference(SysCon), pp. 16, 20 12.
12. Qinlu He, Zhanhuai Li, Xiao Zhang, "Data Deduplication Techniques," Proceedings of IEEE International Conference on Future Information Technology and Management Engineering,Changzhou, China, FITME, pp.430-433, 2010.
13. Peter Shaojui Wang, Feipei Lai, (Senior Member, Ieee), Hsu-Chun Hsiao, And Ja-Ling Wu, (Fellow, Ieee), "Insider Collusion Attack On Privacy-Preserving Kernel-Based Data Mining Systems", IEEE Access, Received April 18, 2016, Accepted April 25, 2016, Date of Publication April 29, 2016, Date of Current Version May 23, 2016,Volume 4, 2016.
14. 2015 Verizon Data Breach Investigations Report, Verizon, Bedminster, NJ, USA, 2015.
15. W. R. Claycomb and A. Nicoll, "Insider threats to cloud computing: Directions for new research challenges," in Proc. IEEE 36th Annu. Comput. Softw. Appl. Conf. (COMPSAC), Jul. 2012, pp. 387_394.
16. Chun-I Fan, Shi-Yuan Huang, Wen-Che Hsu, "Encrypted Data Deduplication in Cloud Storage", 2015 10th Asia Joint Conference on Information Security.