International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 24 Issue 8 – MAY 2017.

AN ENHANCED MULTI-LAYERED CRYPTOSYSTEM BASED SECURE AND AUTHORIZED DEDUPLICATION MODEL IN CLOUD STORAGE SYSTEM

CHEDULURI N VENKATA KIRANKUMAR [#1] and T.RAJENDRA PRASAD [*2]

[#] *PG Scholar, Kakinada Institute Of Engineering & Technology Department of Computer Science & Engineering, JNTUK,A.P, India.*

[* *] *Assistant Prof, Dept of CSE, Kakinada Institute of Engineering & Technology, JNTUK, A.P, INDIA.*

***Abstract*— This paper is an attempt to enhance the Security Model of SecCloud+ with multi-layered cryptosystem based Secure and Authorized Auditing deduplication model. In this paper, we present a scheme that permits a more fine-grained trade-off. The intuition is that outsourced data may require different levels of protection, depending on how popular it is: content shared by many users, such as a popular song or video, arguably requires less protection than a personal document, the copy of a payslip or the draft of an unsubmitted scientific paper. As more corporate and private users outsource their data to cloud storage providers, recent data breach incidents make end-to-end encryption an increasingly prominent requirement. Unfortunately, semantically secure encryption schemes render various cost-effective storage optimization techniques, such as data deduplication, ineffective. We present a novel idea that differentiates data according to their popularity. Based on this idea, we design an encryption scheme that guarantees semantic security for unpopular data and provides weaker security and better storage and bandwidth benefits for popular data. This way, data deduplication can be effective for popular data, whilst semantically secure encryption protects unpopular content. We show that our scheme is secure under the Symmetric External Decisional Diffe-Hellman Assumption in the random oracle model.**

Index Terms*— *De duplication, authorized duplicate check, confidentiality, hybrid cloud

I. INTRODUCTION

Cloud computing technique which is most widely used today's. In that, computing is done over the large communication network like Internet. It is an important solution for business storage in low cost. Cloud computing provide vast storage in all sector like government, enterprise, also for storing our personal data on cloud [1]. Without background implementation details, Platform user can access and share different resources on cloud. The most important problem in cloud computing is that large amount of storage space and security issues. One critical challenge of cloud storage to management of ever-increasing volume of data to improve scalability, storage problem data deduplication is most important technique and has attracted more attention recently [2]. It is an important technique for data compression, It simply avoid the duplicate copies of data and store single copy of data. Data deduplication take place in either block level or file level. In file level approach duplicate files are eliminate, and in block level approach duplicate blocks of data that occur in non-identical files. Deduplication reduces the storage needs by up to 90-95% for backup application, 68% in standard file system. Important issues in data deduplication that security and privacy to protect the data from insider or outsider attack [3].

Presently cloud service provide to the users accessible high available storage and particularly parallel computing of resources at comparatively with different privileges store data on cloud is a most brave issue in organization cloud data storage system [7]. Deduplication is methods which make data manage more scalable in cloud computing [2]. Data deduplication describes as data compression method which eradicates second copy of repeat data in storage space. This method is use to progress storage utilization and also affect to decrease the number of bytes that must be sent before upload in data transmit. In its place to keep same satisfied data copies multiple times deduplication eliminate repetitive data and keep only one physical copy whereas submit other particular unnecessary data to that copy [3]. Deduplication can be applied to data which are in major storage, cloud storage, backup storage for replication transfers [1]. Mostly 3 types are in consideration which are as perfect deduplication process type as block level, second is file level and third is byte level by the names itself deduplication process worked respectively on that content. Users with confidential data are worried about both outsider/insider attacks. So deduplication of data must be hold safety and privacy. But with conventional encryption dissimilar users encrypt data with their own key, which makes similar data with dissimilar user key makes different ciphertext for that data which is not capable for deduplication. The convergent encryption allows encrypt/decrypt data with convergent key on the data thus makes achievable to relate to check duplicates [3]. Therefore with uploading user's data as ciphertext to cloud determined security issues. In order to stop the un authorized access

proofs of ownership protocol can be used as privacy constraint [4]. For data confidentiality, encryption is used by different user for encrypt their files or data, using a secrete key user perform encryption and decryption operation. For uploading file to cloud user first generate convergent key, encryption of file then load file to the cloud. To prevent unauthorized access proof of ownership protocol is used to provide proof that the user indeed owns the same file when deduplication found. After the proof, server provides a pointer to subsequent user for accessing same file without needing to upload same file. When user want to download file he simply download encrypted file from cloud and decrypt this file using convergent key [5].

In this Proof of ownership user can download the decrypted and acquire exact data with convergent keys by specifying its ownership. Therefore by using convergent encryption and proof of ownership both safety and privacy issues determine. At rest the scheme can't effort on privilege level field, it means user can upload file with some set of permissions on its and on the basis on convergent encryption doesn't offer any deduplication on it [1]. As a result it will not support duplicate check with different privileges set provided by the data owner.

In this paper, aiming at efficiently solving the problem of deduplication with differential privileges in cloud computing, we consider a hybrid cloud architecture consisting of a public cloud and a private cloud. Unlike existing data deduplication systems, the private cloud is involved as a proxy to allow data owner/users to securely perform duplicate check with differential privileges. Such architecture is practical and has attracted much attention from researchers. The data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud. A new deduplication system supporting differential duplicate check is proposed under this hybrid cloud architecture where the SCSP resides in the public cloud. The user is only allowed to perform the duplicate check for files marked with the corresponding privileges. Furthermore, we enhance our system in security. Specifically, we present an advanced scheme to support stronger security by encrypting the file with differential privilege keys. In this way, the users without corresponding privileges cannot perform the duplicate check. Furthermore, such unauthorized users cannot decrypt the ciphertext even collude with the S-CSP. Security analysis demonstrates that our system is secure in terms of the definitions specified in the proposed security model. Finally, we implement a prototype of the proposed authorized duplicate check and conduct test bed experiments to evaluate the overhead of the prototype. We show that the overhead is minimal compared to the normal convergent encryption and file upload operations.

II. LITERATURE SURVEY

Some existing works in this area is as follows:

Hybrid Cloud is the architecture that provides the Organization to efficiently work on both the private and public cloud architecture in combination by providing the scalability to adopt. Here some of the basic concepts and idea proposed by authors and how best and easy to adopt this environment is explained by Neal Leavitt. [3] An intelligent workload factoring, service for organization customers which makes the best use of the present public Cloud services including their private owned data centers. It allows the organization to work between the off-premises and the on-premises infrastructure. The efficient core technology that is used for intelligent workload factoring is a fast redundant data element detection algorithm, that helps us factoring all the incoming requests based on the data content and not only on volume of data, Hui Zhang, Guofei Jiang, Kenji Yoshihira, Haifeng Chen and AkhileshSaxena. [4]

The term ―Cloud has many definitions one among them is to provide infrastructure as a service system where the IT infrastructure will be deployed in the particular cloud service provider, data center as virtual machine. The growing popularity of laas will help us to transform the organization present infrastructure into the required hybrid cloud or private cloud. Open Nebula Concept is being used that will provide the features that are not present in any other cloud software, Borja Sotomayor ,Rubén S. Montero and Ignacio M. Llorente, Ian Foster. [5] Data Deduplication is a technique that is mainly used for reducing the redundant data in the storage system which will unnecessarily use more bandwidth and network. So here some common technique is being defined which finds the hash for the particular file and with that the process of deduplication can be simplified, David Geer. [6] De-duplication is the technique that is most effective most widely used but when it is applied across the multiple users the cross-user deduplication tend to have to many serious privacy implications. Simple mechanisms can be used which can enable the cross-user deduplication which will reduce the risks of the data leakage and also some of the security issues are discussed with how exactly to identify the files and to encrypt them while sending is discussed, Danny Harnik, Benny Pinkas, Alexandra Shulman- Peleg. [7]

M. Bellare [8] design a system, DupLESS that combines a CE-type scheme with the ability to obtain message-derived keys with the help of a key server (KS) shared amongst a group of clients. The clients interact with the KS by a protocol for oblivious PRFs, ensuring that the KS can cryptographically mix in secret material to the per message keys while learning nothing about files stored by clients. These mechanisms ensure that Dup LESS provides strong security against external attacks and that the security of DupLESS gracefully degrades in the face of comprised systems. Should a client be compromised, learning the plaintext underlying another client's cipher text requires mounting an online brute force attacks. Aim of M. Bellare [9] is to formalize a new cryptographic primitive, Message-Locked Encryption (MLE), 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. They provide definitions both for privacy and for a form of integrity that they call tag consistency. They provide ROM security analyses of a natural family of MLE schemes that includes deployed schemes. They make connections with deterministic encryption, hash functions secure on correlated inputs. G. Neven [10] provides

either security proofs or attacks for a large number of identity-based identification and signature schemes defined either explicitly or implicitly in existing literature. Underlying these is a framework that on the one hand helps explain how these schemes are derived and on the other hand enables modular security analyses, thereby helping to understand, simplify, and unify previous work. They also analyze a generic folklore construction that in particular yields identity-based identification and signature schemes without random oracles.

J. Xu [11] proposed growing need for secure cloud storage services and the attractive properties of the convergent cryptography lead us to combine them, thus, defining an innovative solution to the data outsourcing security and efficiency issues. Our solution is based on a cryptographic usage of symmetric encryption used for enciphering the data file and asymmetric encryption for meta data files, due to the highest sensibility of these information towards several intrusions. In addition, thanks to the Merkle tree properties, this proposal is shown to support data deduplication, as it employs an pre-verification of data existence, in cloud servers, which is useful for saving bandwidth. Besides, our solution is also shown to be resistant to unauthorized access to data and to any data disclosure during sharing process, providing two levels of access control verification. Finally, we believe that cloud data storage security is still full of challenges and of paramount importance, and many research problems remain to be identified.

III. EXISTING SYSTEM

Data deduplication systems, the private cloud is involved as a proxy to allow data owner/users to securely perform duplicate check with differential privileges. Such architecture is practical and has attacked so much attention from researchers. The data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud.

A. Issues in Existing System

1. Traditional encryption, while providing data confidentiality, is incompatible with data deduplication.
2. Identical data copies of different users will lead to different cipher texts, making deduplication impossible.

IV. PROPOSED SYSTEM

In this paper, we enhance our system in security. Specifically, we present an advanced scheme to support stronger security by encrypting the file with differential privilege keys. Security analysis demonstrates that our system is secure in terms of the definitions specified in the proposed security model. In our system we implement a project that includes the public cloud and the private cloud and also the hybrid cloud which is a combination of the both public cloud and private cloud. In general by if we used the public cloud we can't provide the security to our private data and hence our private data will be loss. So that we have to provide the security to our data for that we make a use of private cloud also. When we use a private cloud the greater security can be provided. In this system we also provide the data deduplication. Which is used to avoid the duplicate copies of data, User can upload and download the files from public cloud but private cloud provides the security for that data. That means only the authorized person can upload and download the files from the public cloud. For that user generates the key and stored that key onto the private cloud. At the time of downloading user request to the private cloud for key and then access that Particular file.

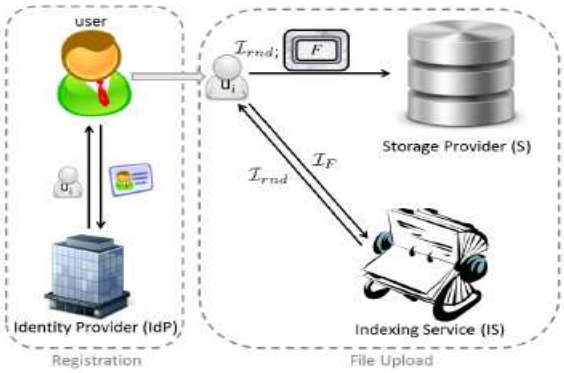

Fig.1 System Architecture

A. SYSTEM MODEL

1) Hybrid Architecture for Secure Deduplication

At a high level, our setting of interest is an enterprise network, consisting of a group of affiliated clients (for example, employees of a company) who will use the S-CSP and store data with deduplication technique. In this setting, deduplication can be frequently used in these settings for data backup and disaster recovery applications while greatly reducing storage space. Such systems are widespread and are often more suitable to user file backup and synchronization applications than richer storage abstractions.

a) PRIVATE CLOUD

Compared with the traditional deduplication architecture in cloud computing, this is a new entity introduced for facilitating user's secure usage of cloud service. Specifically, since the computing resources at data user/owner side are restricted and the public cloud is not fully trusted in practice, private cloud is able to provide data user/owner with an execution environment and infrastructure working as an interface between user and the public cloud. The private keys for the privileges are managed by the private cloud, who answers the file token requests from the users. The interface offered by the private cloud allows user to submit files and queries to be securely stored and computed respectively.

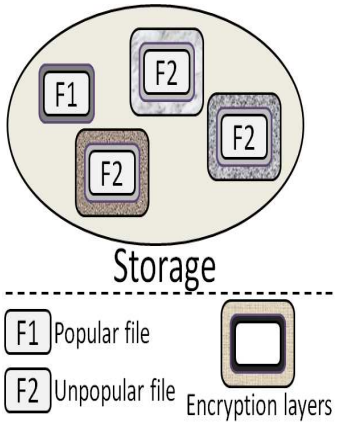

Fig.2 Proposed System

Notice that his is a novel architecture for data deduplication in cloud computing, which consists of a twin clouds (i.e., the public cloud and the private cloud).Actually, this hybrid cloud setting has attracted more and more attention recently. For example, an enterprise might use a public cloud service, such as Amazon S3, for archived data, but continue to maintain in-house storage for operational customer data. Alternatively, the trusted private cloud could be a cluster of virtualized cryptographic co-processors, which are offered as a service by a third party and provide the necessary hardware based security features to implement a remote execution environment trusted by the users.

2) ADVERSARY MODEL

Typically, we assume that the public cloud and private cloud are both "honest-but-curious". Specifically they will follow our proposed protocol, but try to find out as much secret information as possible based on their possessions. Users would try to aces data either within or out of the scopes of their privileges. In this paper, we suppose that the files are sensitive and needed to be fully protected against both public cloud and private cloud. Under the assumption, two kinds of adversaries are considered, that is,

1)external adversaries which aim to extract secret information as much as possible from both public cloud and private cloud;

2) internal adversaries who aim to obtain more information on the file from the public cloud and duplicate-check token information from the private cloud outside of their scopes. Such adversaries may include S- CSP, private cloud server and authorized users. The detailed security definitions against these adversaries are discussed, where attacks launched by external adversaries are viewed as special attacks from internal adversaries.

3) DESIGN GOALS

In this paper, we address the problem of privacy preserving deduplication in cloud computing and propose a new deduplication system supporting for cloud

a) DIFFERENTIAL AUTHORIZATION

Each authorized user is able to get his/her individual token of his file to perform duplicate check based on his privileges. Under this assumption, any user cannot generate a token for duplicate check out of his privileges or without the aid from the private cloud server.

b) AUTHORIZED DUPLICATE CHECK

Authorized user is able to use his/her individual private keys to generate query for certain file and the privileges he/she owned with the help of private cloud, while the public cloud performs duplicate check directly and explains the user if there is any duplicate. The security requirements considered in this paper lie in two folds, including the security of file token and security of data files. For the security of file token, two aspects are defined as UN forge ability and in distinguish ability of file token. The details are given below.

c) UNFORGEABILITY OF FILE TOKEN/DUPLICATE-CHECK TOKEN

Unauthorized users without appropriate privileges or file should be prevented from getting or generating the file tokens for duplicate check of any file stored at the S-CSP. The users are not allowed to collude with the public cloud server to break the UN forge ability of file tokens. In our system, the S-CSP is honest but curious and will honestly perform the duplicate check upon receiving the duplicate request from users. The duplicate check token of users should be issued from the private cloud server in our scheme.

d) INDISTINGUISHABLY OF FILE TOKEN/DUPLICATE-CHECK TOKEN

It requires that any user without querying the private cloud server for some file token, he cannot get any useful information from the token, which includes the file information or the privilege information.

e) DATA CONFIDENTIALITY

Unauthorized users without appropriate privileges or files, including the S-CSP and the private cloud server, should be prevented from aces to the underlying plaintext stored at S-CSP. In another word, the goal of the adversary is to retrieve and recover the files that do not belong to them. In our system, compared to the previous definition of data confidentiality based on convergent encryption, a higher level confidentiality is defined and achieved.

V. IMPLEMENTATION

The proposed system of this project is divided into four major modules and described as below.

1. CLOUD SERVICE PROVIDER
2. DATA USERS MODULE
3. PRIVATE CLOUD MODULE
4. SECURE DEDUPLICATION SYSTEM

A. MODULES DESCRIPTION

1) CLOUD SERVICE PROVIDER

In this module, we develop Cloud Service Provider module. This is an entity that provides a data storage service in public cloud. The S-CSP provides the data outsourcing service and stores data on behalf of the users.

To reduce the storage cost, the S-CSP eliminates the storage of redundant data via de duplication and keeps only unique data. In this paper, we assume that S-CSP is always online and has abundant storage capacity and computation power.

2) DATA USERS MODULE

A user is an entity that wants to outsource data storage to the S-CSP and aces the data later. In a storage system supporting de duplication, the user only uploads unique data but does not upload any duplicate data to save the upload bandwidth, which may be owned by the same user or different users. In the authorized de duplication system, each user is issued a set of privileges in the setup of the system. Each file is protected with the convergent encryption key and privilege keys to realize the authorized de duplication with differential privileges

3) PRIVATE CLOUD MODULE

Compared with the traditional de duplication architecture in cloud computing, this is a new entity introduced for facilitating user's secure usage of cloud service. Specifically, since the computing resources at data user/owner side are restricted and the public cloud is not fully trusted in practice, private cloud is able to provide data user/owner with an execution environment and infrastructure working as an interface between user and the public cloud. The private keys for the privileges are managed by the private cloud, who answers the file token requests from the users. The interface offered by the private cloud allows user to submit files and queries to be securely stored and computed respectively

4) SECURE DEDUPLICATION SYSTEM

We consider several types of privacy we need protect, that is, i) UN forge ability of duplicate- check token: There are two types of adversaries, that is, external adversary and internal adversary. As shown below, the external adversary can be viewed as an internal adversary without any privilege. If a user has privilege p, it requires that the adversary cannot forge and output a valid duplicate token with any other privilege p′ on any file F, where p does not match p′. Furthermore, it also requires that if the adversary does not make a request of token with its own privilege from private cloud server, it cannot forge and output a valid duplicate token with p on any F that has been queried.

Fig.3 File upload to auditor

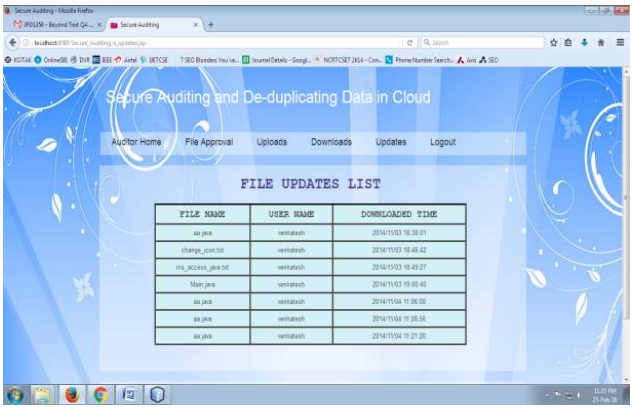

Fig.4 File updated list

Though the above solution supports the differential privilege duplicate, it is inherently subject to brute force attacks launched by the public cloud server, which can recover files falling into a known set. Will be insecure for predictable file. We design and implement a new system which could protect the security for predictable message. The main idea of our technique is that he novel encryption key generation algorithm. More specifically, Security is thus only possible when such a message is unpredictable. This traditional convergent encryption for simplicity, we will use the hash functions to define the tag generation functions and convergent keys in this section. In traditional convergent encryption, to support duplicate check, the key is derived from the file F by using some cryptographic hash function.

VI. CONCLUSION

In this paper, the notion of authorized data deduplication was proposed of users in the duplicate check. We also presented several to protect the data security by including differential privileges new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture, in which the duplicate-check tokens of files are generated by the private cloud schemes are secure in terms of insider and outsider server with private keys. Security analysis demonstrates that our attacks specified in the proposed security model of our proposed authorized duplicate check scheme and conduct test bed experiments on our prototype. As a proof of concept, we implemented a prototype we showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent

encryption and network.

REFERENCES

[1] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou, "A Hybrid Cloud Approach for Secure Authorized Deduplication", IEEE Transactions on Parallel and Distributed Systems, 2014.
[2] Open SSL Project. http://www.openssl.org/.
[3] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.
[4] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicates storage. In USENIX Security Symposium, 2013.
[5] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message locked encryption and secure deduplication. In EUROCRYPT, pages 296–312, 2013.
[6] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 22(1):1–61, 2009.
[7] M. Bellare and A. Palacio. Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks. In CRYPTO, pages 162–177, 2002.
[8] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. Twin clouds: An architecture for secure cloud computing. In Workshop on Cryptography and Security in Clouds (WCSC 2011), 2011.
[9] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer. Reclaiming space from duplicate files in a server less distributed file system. In ICDCS, pages 617–624, 2002.
[10] D. Ferraiolo and R. Kuhn. Role-based access controls. In 15th NIST-NCSC National Computer Security Conf., 1992.
[11] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011.
[12] C. Ng and P. Le. Revdedup: A reverse Deduplication Storage system optimized for reads to latest backups. In Proc. of APSYS, Apr 2013.
[13] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S. Osowski and P. Leca, editors, Procedings of the 27th Annual ACM Symposium on Applied Computing, pages 41–46. ACM, 2012.
[14] R. D. Pietro and A. Sornioti. Bosting eficiency and securityin proof of ownership for de duplication. In H. Y.Youm andY. Won, editors, ACM Symposium on Information, Computer and Communications Security, pages 81–82. ACM, 2012.
[15] S. Quinlan and S. Dorward.Venti: a new approach to Archival storage. In Proc. USENIX FAST, Jan 2002.
[16] A. Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Le, and J. C. S. Lui. A secure cloud backup system with assured deletion and version control. In 3rd International Workshop on Security in Cloud Computing, 2011.

AUTHOR PROFILE

CHEDULURI N VENKATA KIRANKUMAR is a student of Kakinada Institute Of Engineering & Technology affiliated to JNTUK, Kakinada pursuing M.Tech (Computer Science & Engineering). His Area of interest includes Cloud Computing and its objectives in all current trends and techniques in Computer Science.

T. RAJENDRA PRASAD M.TECH is working as Assistant Professor, Department of Computer Science & Engineering, Kakinada Institute of Engineering & Technology, JNTUK, A.P, INDIA.