Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 3, March 2016, pg.330 – 335

AN EFFICIENT TECHNIQUE FOR DATA COMPRESSION AND CONVERGENT ENCRYPTION IN THE HYBRID CLOUD

[1]Mr. D.Jaya Kumar, [2]Ms. V.Deepika, [3]Ms. A.Ishwarya, [4]Ms. N.Monisha

[1]Department of Information Technology, Prathyusha Engineering College, Thiruvallur, India
[2]Department of Information Technology, Prathyusha Engineering College, Thiruvallur, India
[3]Department of Information Technology, Prathyusha Engineering College, Thiruvallur, India
[4]Department of Information Technology, Prathyusha Engineering College, Thiruvallur, India
[1] jkjayakumard2010@gmail.com, [2] raapallideepika@gmail.com, [3] ishwarya795@gmail.com, [4] yashvie475@gmail.com

ABSTRACT: In cloud computing, there is an enormous amount of data will be stored. So there is a need for high security and a technique to manage the data storage. For that, we are providing deduplication that is a specialized data compression method for avoidance of redundant data storage in the cloud. An another technique for security is a convergent encryption technique. It is a content hash keying method that provides confidentiality of data before outsourcing it. Here we implement an enhanced process in deduplication where it has to done before outsourcing into the cloud. This will minimize the overhead and also reduce the volume of data in the cloud.

Keywords: Deduplication, convergent encryption, confidentiality, hybrid cloud

I. INTRODUCTION

Cloud computing means "a type of Internet-based computing," where different services — such as servers, storage and applications —are delivered to an organization's computers and devices through the Internet. The major problem in cloud is data storage and it management. Cloud computing uses networks of large groups of servers typically running low-cost consumer PC technology with specialized connections to spread data-processing chores across them. This shared IT infrastructure contains large pools of systems that are linked together. Often, virtualization techniques are used to maximize the power of cloud computing. Data Deduplication [7] is one of the storage data compression technique that enables the avoidance of redundant data in a file. Data deduplication is done in block or file level. At block level, the block of data in non identical files will be compared and eliminates the replicated copy.

The intent to provide security of data, encryption can be done. In traditional encryption, the identical content with different user keys will produce the different cipher text. So that we are using convergent encryption [4] technique which provides a same cipher text even with different user key. After the generation of key and encrypted data, cipher text will be send to the cloud.

While uploading a file, the user with differential privileges will undergoes duplicate check. When a subsequent user uploads the same file, then a pointer will be maintained in the server instead of storing it again in the cloud. The encrypted file can be downloaded by the corresponding data owners by using their convergent keys.

In the previous paper, the private cloud is involved as a proxy to allow data users/owners to securely perform duplicate check with differential privileges and the data will be stored in the public cloud. To avoid the repetition of data in the public cloud, deduplication has been done inside the cloud and to protect the confidentiality of sensitive data while supporting deduplication, cloud computing provides unlimited resources to the user. Cloud service provides storage and parallel computing resources a relatively low cost. The data stored in the cloud and shared by user with the certain privileges. One critical challenge of cloud storage services is the management of the ever - increasing volume of data. Increased processing overhead inside the cloud. In this paper, we enhance the security by providing convergent encryption technique which is compatible with data deduplication. The stronger security has been provided by encrypting the file with differential privilege keys .The convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security , this paper makes the first attempt to formally address the problem of authorized data deduplication . Different from traditional deduplication system check besides the data itself. We also present several new deduplication constructions supporting authorized duplicate check before storing in a hybrid cloud architecture. The management of the ever - increasing volume of data has done efficiently. Reduce the storage processing inside the cloud and incurs the minimal overhead than normal operation. Reduce the storage size of the tags for integrity check and enhance data confidentiality.

II. LITERATURE SURVEY

A. A Secure Cloud Backup System with Assured Deletion and Version Control

Cloud storage is an emerging service model that enables individuals and enterprises to outsource the storage of data backups to remote cloud providers at a low cost. However, cloud clients must enforce security guarantees of their outsourced data backups. We present *FadeVersion*, a secure cloud backup system that serves as a security layer on top of today's cloud storage services. FadeVersion follows the standard version-controlled backup design, which eliminates the storage of redundant data across different versions of backups. On top of this, FadeVersion applies cryptographic protection to data backups. Specifically, it enables fine-grained assured deletion, that is, cloud clients can assuredly delete particular backup versions or files on the cloud and make them permanently inaccessible to anyone, while other versions that share the common data of the deleted versions or files will remain unaffected. We implement a proof-of-concept prototype of FadeVersion and conduct empirical evaluation a top Amazon S3. We show that FadeVersion only adds minimal performance overhead over a traditional cloud backup service that does not support assured deletion**.**

B. Private Data Deduplication Protocols in Cloud Storage

In this paper, a new notion which we call private data Deduplication protocol, a deduplication technique for private data storage is introduced and formalized. Intuitively, a private data deduplication protocol allows a client who holds a private data proves to a server who holds a summary string of the data that he/she is the owner of that data without revealing further information to the server. Our notion can be viewed as a complement of the state-of-the-art public data deduplication protocols of Halevi et al . The security of private data deduplication protocols is formalized in the simulation-based framework in the context of two-party computations. A construction of private deduplication protocols based on the standard cryptographic assumptions is then presented and analyzed. We show that the proposed private data deduplication protocol is provably secure assuming that the underlying hash function is collision-resilient, the discrete logarithm is hard and the erasure coding algorithm can erasure up to _-fraction of the bits in the presence of malicious adversaries in the presence ofmalicious adversaries. To the best our knowledge this is the first deduplication protocol for private data storage.

C. DupLESS: Server-Aided Encryption for Deduplicated Storage

Cloud storage service providers such as Dropbox,Mozy,and others perform deduplication to save space by only storing one copy of each file uploaded. Should clients conventionally encrypt their files, however, savings are lost. Message-locked encryption (the most prominent manifestation of which is convergent encryption) resolves this tension. However it is inherently subject to brute-force attacks that can recover files falling into a known set. We propose an architecture that provides secure deduplicated storage resisting brute-force attacks, and realize it in a system called DupLESS. In DupLESS, clients encrypt under message-based keys obtained from a key-server via an oblivious PRF protocol. It enables clients to store encrypted data with an existing service, have the service perform deduplication on their behalf, and yet achieves strong confidentiality guarantees. We show that encryption for deduplicated storage can achieve performance and space savings close to that of using the storage service with plaintext data.

D. Reclaiming Space from Duplicate Files in a Serverless Distributed File System

The Farsite distributed file system provides availability by replicating each file onto multiple desktop computers. Since this replication consumes significant storage space, it is important to reclaim used space where possible. Measurement of over 500 desktop file systems shows that nearly half of all consumed space is occupied by duplicate files. We present a mechanism to reclaim space from this incidental duplication to make it available for controlled file replication. Our mechanism includes 1)convergent encryption, which enables duplicate files to coalesced into the space of a single file, even if the files are encrypted with different users' keys, and 2) SALAD, a Self-Arranging, Lossy, Associative Database for aggregating file content and location information in a decentralized, scalable, fault-tolerant manner. Large-scale simulation experiments show that the duplicate-file coalescing system is scalable, highly effective, and fault-tolerant.

E. Security Proofs for Identity-Based Identification and Signature Schemes

This paper 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. We also analyze a generic folklore construction that in particular yields identity-based identification and signature schemes without random oracles

III. 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. In this way, the users without corresponding privileges cannot perform the duplicate check. Furthermore, such unauthorized users cannot decrypt the cipher text 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.

A. ADVANTAGES OF PROPOSED SYSTEM:

- The user is only allowed to perform the duplicate check for files marked with the corresponding privileges.
- We present an advanced scheme to support stronger security by encrypting the file with differential privilege keys.
- Reduce the storage size of the tags for integrity check. To enhance the security of deduplication and protect the data confidentiality,

IV. ENCRYPTION SCHEMES

In this section, we have includes some primary encryption schemes in cryptography

A. Symmetric Encryption

A symmetric encryption is a common method which uses a secret key k to encrypt and decrypt a file.

A symmetric encryption scheme consists of three primitive functions:

- KeyGen→ k is the key generation algorithm that generates k using security parameter.
- Encrypt(k,M) → C is the symmetric encryption algorithm that takes the secret k and message M and then outputs the ciphertext C.
- Decrypt(k,C)→ M is the symmetric decryption algorithm that takes the secret and ciphertext C and then outputs the original message M.

B. Convergent Encryption

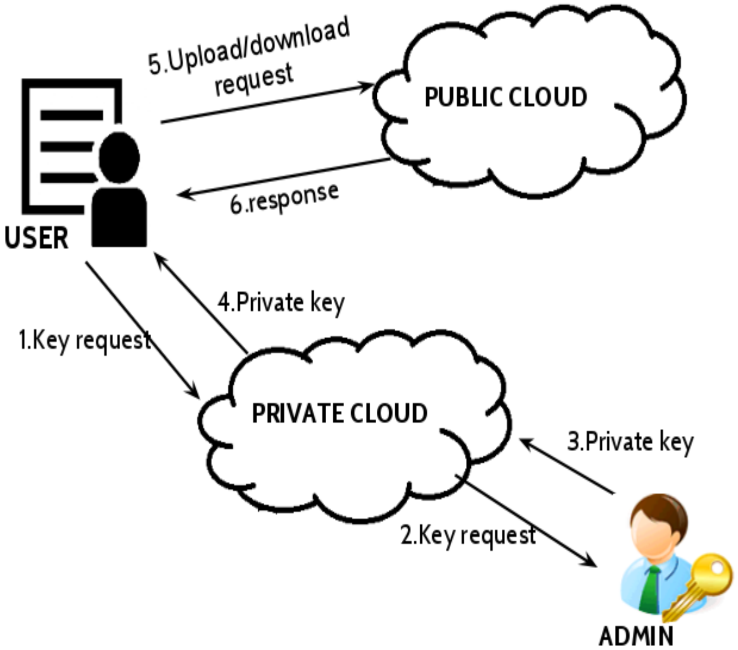
Convergent encryption [8] , [4] is a technique that supports data deduplication. In this technique, if the user different keys that encrypts a identical plain text file will produce a same cipher text. So that while comparing the file for duplicate check it will show that the content is same.

The basic idea for convergent encryption is as follows:

The key K=H(M) where M is the message that need to be encrypt and H is a hash function. Then the encrypted message file will be C=E(K;M)=E(H(M);M).

V. SYSTEM ARCHITECTURE

In the high level enterprise network, the deduplication with S-CSP can be used in the cloud. Deduplication can be use for recovery of data in disaster and also data backup. There are three entities in this system- user, public cloud and private cloud. Data deduplication was performed in S-CSP public cloud.



The initial request for the file will be given in the private cloud as a token. The duplicate check has been done in file level. If the files are identical, then obviously there blocks will be same. The block level check will be done similar to file level deduplication [9].

The modules that are available in the system architecture are

- ❖ Cloud Service Provider
- ❖ Data Users Module

- Private Cloud Module
- Secure Deduplication System

A. Cloud Service Provider

Public clouds are owned and operated by companies that offer rapid access over a public network to affordable computing resources. With public cloud services, users don't need to purchase hardware, software or supporting infrastructure, which is owned and managed by providers. In Cloud Service Provider module, it acts as an entity which will provides a data storage service in public cloud. The S-CSP(Storage-Cloud Service Provider) provides the data outsourcing service and stores data on behalf of the users. The S-CSP eliminates the storage of redundant data and stores only the unique data and that will reduce the data cost. In this paper, we assume that S-CSP is always online and has abundant storage capacity and computation power.

B. Data Users Module

A user is an entity who wants to outsource data to the S-CSP and that data can be accessed later. User have to register in the cloud. User will login into the cloud and then he/she will be provided with a key. In a storage system supporting deduplication, unique data alone allowed in the cloud. In the authorized deduplication system, each file is protected with the convergent encryption key and privilege keys to realize the authorized deduplication with differential privileges.

C. Private Cloud Module

A private cloud is infrastructure operated solely for a single organization, whether managed internally or by a third party, and hosted either internally or externally. Private clouds can take advantage of cloud's efficiencies, while providing more control of resources and steering clear of multi-tenancy. Private cloud module is a new entity introduced to facilitate secured usage of cloud service. The private cloud acts as an intermediate between the users and the public cloud. Since the public cloud is not fully trusted, the private cloud provides and manages the key for secured access of data in the public cloud. the private cloud is the one who respond for 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.

D. Secure Deduplication System

Here we have considered the unforgeability of duplicate-check . The adversaries is of two types, that is, external adversary and internal adversary. The external adversary can be viewed as an internal adversary without any privilege. The internal adversary is the one who has privilege to access the cloud and tries to access the data of unauthorized privileges. If a user has 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.

VI. CONCLUSION

In this paper, the idea of authorized data deduplication was proposed to protect the data security by including differential authority of users in the duplicate check. In public cloud our data are securely store in encrypted format and in private cloud our key is store with respective file. There is no need for the user to remember the key and without key nobody can access our file or data which is present public cloud. In our paper, we also done the duplicate check outside the public cloud so that it minimize the overhead and which inturns increases the performance inside the cloud.

REFERENCES

[1] D.JayaKumar, A Resourceful allocation of data storage in cloud computing In *International Journal of Advance in Engineering and Engineering Technology*,2013

[2] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Server aided encryption for deduplicated storage. In *USENIX Security Symposium*, 2013.

[3] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplicationprotocols in cloud storage. In S. Ossowski and P. Lecca, editors,*Proceedings of the 27th Annual ACM Symposium on Applied Computing*,pages 441–446. ACM, 2012.

[4] Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, 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.
[5] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer.Reclaiming space from duplicate files in a serverless distributed file system. In *ICDCS*, pages 617–624, 2002.
[6] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity based identification and signature schemes. *J. Cryptology*,22(1):1–61, 2004.
[7] J.Stanek,A.Sorniotti,E.Androulaki, and L. Kencl. A secure data deduplication scheme for cloud storage. In *Technical Report*, 2013.
[8] S. Quinlan and S. Dorward, "Venti: A new approach to archival storage," in Proc. 1st USENIX Conf. File Storage Technol., Jan. 2002, p.7.
[9] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," in Proc. 32nd Annu. Int. Conf. Theory Appl. Cryptographic Techn., 2013, pp. 296–312.
[10] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, "Secure dedu-plication with efficient and reliable convergent key management," in Proc. IEEE Trans. Parallel Distrib. Syst., http:// doi.ieeecomputersociety.org/10.1109/TPDS.2013.284, 2013.