International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

A Secure Authorized Hybrid Cloud Distributed Key Generation for Encrypted Deduplication of Data

Akanksha V. Patil[1], Navnath D. Kale[2]

[1]Student in Department of Computer Engineering, P.V.P.I.T., Bavdhan, Pune, Maharashtra, India

[2]Assistant Professor, Department of Computer Engineering, P.V.P.I.T., Bavdhan, Pune, Maharashtra, India

Abstract: ***In recent days, cloud computing became an emerging service model which provides highly available storage and massively parallel computing. Cloud storage enables users to out- source their data backup over remote cloud providers. Managing ever increasing growth of data, is became a great headache. To address this problem, data deduplication technique is introduced which eliminates redundant data copies by keeping a physical copy. For security concerns encryption becomes a necessary before updating data into the cloud. Since these are two challenges that we focused in this paper. For achieving deduplication along with data security, secure hashing algorithm is used.***

Keywords: cloud computing; cloud storage; de- duplication; encryption;

1. Introduction

Cloud computing is internet-based computing where large groups of remote servers are connected to each other to allow the centralized data storage, and online access to computer services or resources. There are three types of cloud - public cloud, private cloud and hybrid cloud. In public cloud, applications and storage are available over the internet for general use. In private cloud, a virtualized data center is used that operates within a firewall. In this research introduce hybrid cloud which is mix of public and private cloud.

Cloud computing focused on maximizing the effectiveness of the shared resources. It provides computation and storage resources on the Internet. Cloud resources are usually shared by multiple users and also it is dynamically reallocated per demand. By using cloud computing, many numbers of users can access a single server to retrieve and update their data without purchasing licenses for different applications.

Exponential growth of ever increasing data over cloud is became a critical challenge. To face that challenge data deduplication technique is introduced. Data deduplication is data compression technique which eliminate repeated data. Instead of taking multiple numbers of copies of same data, it saves just one copy of the data and other copies are replaced with pointers that lead back to the original copy. It improves bandwidth efficiency and storage utilization.

Data dedupliaction protects confidentiality of data. Data deduplication work with convergent encryption technique to encrypt the data before uploading, and we enhance some hashing algorithm which makes the technique very secure before uploading encrypted file into the cloud.

2. Literature Survey

In [1], Jin Li proposes deduplication technique. In that he used convergent encryption technique to encrypt data before outsourcing. He also presents authorized duplicate check to provide better data security in cloud. In [2], Sadaqat Ur Rehman investigates different cryptographic techniques. Also he compares different encryption techniques such as stream cipher, block cipher and hashing techniques.

In [3], E.Mounika describes about information deduplica- tion. In that she merged encryption system to encode the in- formation before outsourcing. To better secure information se- curity, she endeavor to formally address the issue of approved information deduplication. In [4], Yinjin Fu proposed AA- Dedupe, an application aware source deduplication scheme, to significantly reduce the computational overhead, increase the deduplication throughput and improve the data transfer efficiency.

In [4], Hugo Krawczyk presents Key derivation system. He provides detailed rationale for the design of KDFs based on the extract-then-expand approach. He presents the general and rigorous dentition of KDFs and their security that base on the notion of computation. He specifies a concrete fully practical KDF based on the HMAC construction, and also it provides an analysis of construction based on the pseudorandom and extraction properties of HMAC and SHA.

3. Background and Related Work

In existing system, process key generation and data encryption, are involved users can retain the keys and send the cipher text to the cloud. Because of the encryption operation is deterministic and it is derived from the data content, identical data copies will generate the same convergent key and hence the same cipher text. To prevent unauthorized access, a secure proof of ownership protocol

is also needed to provide the proof as the user indeed owns the same file when a duplicate is found. After that proof, subsequent users with the same file will provided a pointer from the server without needing to upload the same file. If the user wants to download the encrypted file, then he can download it with the pointer from the server. That file can only be decrypted by the corresponding data owners with their convergent keys. Thus, convergent encryption allows the cloud to perform deduplication on the cipher texts and the proof of ownership prevents the unauthorized user to access the file In such an authorized deduplication system, each user is issued a set of privileges during system we elaborate the definition of a privilege with examples).Each file uploaded to the cloud is also bounded by a set of privileges to specify what kind of users is allowed to perform the duplicate check and also access the files. Before submitting his duplicate check request for some file to the server, the user needs to take this file and his own privileges as inputs.

In proposed system, data deduplication is an important technique for eliminating redundant data. Instead of storing multiple numbers of same files, it stores only one copy of file. In most organizations, storage system contains multiple pieces of duplicate data. For example, various users can save same file in different places. Deduplication eliminates these extra copies by saving just a single copy of the data and replacing the other copies with pointers that lead back to the original copy. It is data compression technique which improves the bandwidth efficiency and also storage utilization. Data deduplication is widely used in cloud computing. It makes data management scalable and handles storage problem in cloud. Data deduplication protects the confidentiality of sensitive data. Data deduplication work with convergent encryption technique to encrypt the data before uploading, and we enhance some hashing algorithm which makes the technique very secure before uploading encrypted file into the cloud. SHA-1 is a most commonly used from SHA series of cryptographic hash functions. SHA-1 can produce a message digest. Normally the input data is often called the message, and the hash value is often called the message digest or simply the digest. A message digest is serves as a means of reliably identifying a file. Each hashing function forms a unique key depend on the file size and content produce in the file, even if there is a slight changes in the file the key for the file will be changed completely for the whole file.

4. Mathematical Model

A. Authentication

Suppose the data owner wants to upload a file, the owner must be privileged user.
Input= User, Data owner, Private cloud, key
Authentication involves following process:

1) User must be a privileged one
2) He generates a key which he can use that for Decryption, another kind of authentication
3) The generated key will be stored on the private cloud

B. Key Generation

A unique key is generated for each file which helps to identify the file duplication
Process= user, file data,
key

1) Key generation involves following process. storage inside the public cloud, key generation generate a
2) User inputs a file. unique key which differ to each file. For retrieving data, user
3) A unique is generated for each file. can directly download data from cloud storage but only after
4) Key will differs from file to file. specifying the user authentication.

C. Anonymization

Once the file is checked in the cloud ,if the cloud does not the file content, the file will be encrypted before it got uploaded in the file.
Upload= File data, key
It involves following procedure
1) File that user inputed
2) Key which is generated in key generation

D. De-Anonymization

If the user wants to download contents from the cloud user must specify the key and download the file contents of the data.
Download= file data, User Specified key
It involves following procedures
1) Anonymized data
2) User Specified key

5. Implementation Details

A. System Architecture

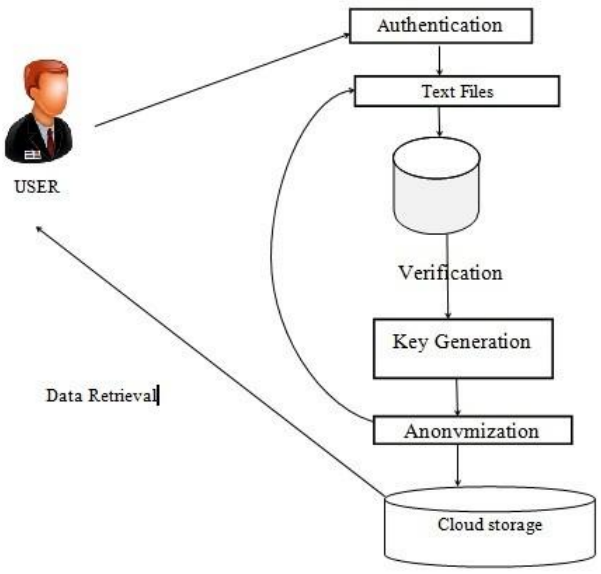

Figure 1: System Architecture

We consider cloud storage system as shown in Fig. 1, which involves data owners (user), the private cloud storage (database), and the public cloud storage (cloud storage). The user firstly get authenticate. Only authenticate user can enter in the system. Authenticate

key is stored in the private cloud. File data is then transferred towards public cloud. For securely

B. System Modules

1) User Authentication: Authentication is accepting proof of identity given by a credible person who has first-hand evidence that the identity is genuine. Attribute comparison might be vulnerable to forgery. In general, it relies on the facts that creating a forgery indistinguishable from a genuine artifact requires expert knowledge, that mistakes are easily made, and that the amount of effort required to forgery is considerably greater than the amount of profit that can be gained from it. Only the privileged user is allowed to store the content on the cloud and allowed to process the further procedures.

2) Key Generation: Key generation is the process of generating key for checking the file duplication which occur on the cloud normally by enabling SHA-1 Algorithm for the key generation, normally individual key will be generated for each file, even when the same file is altered the key for the same file be changed for each individual file and it also very secure than the other algorithm such as HMAC which is used in the proposed system since SHA-1 uses an iterative algorithm. It generates digests by first splitting content into blocks of 64 bytes and, one after the other, combining those blocks together to generate the 20 byte digest.

3) File transmission: A unique key will be generated for each individual file and each file will be checked with cloud whether it is previously exist or not in the cloud if it already present the current file will not be uploading an error signal will be a raised. This will reduce the amount of storage space which occurs on the cloud and also bandwidth will be reduced and also these terms to reduce the cost of cloud whenever user needs to store inside the cloud.

4) Data Anonymization: Data anonymization is normally done to add security inside the cloud whenever the file is not duplicated the file which we want to store inside the cloud will checked with generated key by using SHA-1 algorithm and the file will get encrypted and stored on the cloud.

5) Data De-Anonymization and Downloading: Whenever the user wants its encrypted cloud data to get viewed or downloaded, the user want to specify the key which the user enters during the authentication section once user enters the specified it will checked with the database and file content will be decrypted and it will be downloaded from the cloud.

6. Result

Result of Practical Work:

Figure 2 shows the cloud service provide which acts as cloud server. it has four parts, which are useful for file or block uploading and downloading. CSP give the response for every request of user. in project, it checks duplicate file, also it checks block duplication. it give response as it duplicate or not and also if uploads file in cloud and also downloads it.

Figure 3 shows algorithm comparison. Here in my project, in existing system, author used DES algorithm for block encryption. In proposed system, i used AES algorithm for block encryption. So here in that form, i am going to compare these two algorithms. Here, while encrypting block, DES encrypted block size varies from AES encrypted block size. Which result in the graph in figure 4. Figure 4 shows how AES takes small size than DES. It shows how AES is better than DES. As AES used key to encrypt it, it encrypt a block, which shows too small size than the block is encrypted by DES. So it reduces the storage space and increase storage capacity of overall system. As AES is more secure than DES, it also enhances the security inside the cloud. As on result, we can say that proposed give better result than existing system.

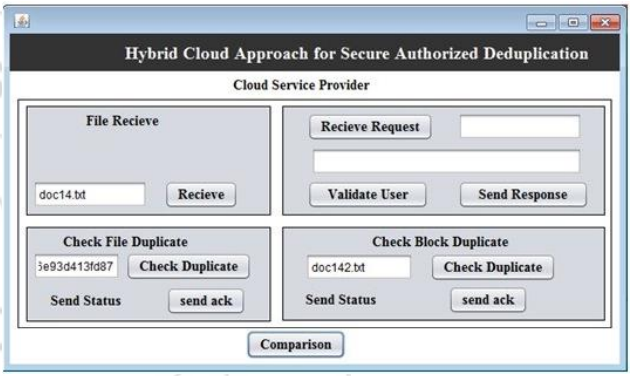

Figure 2: Cloud Service Provider

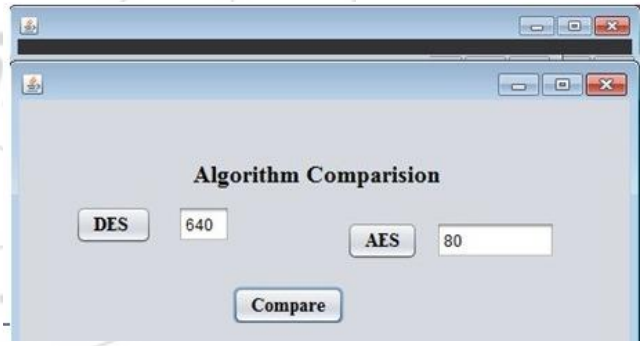

Figure 3: Algorithm Comparison

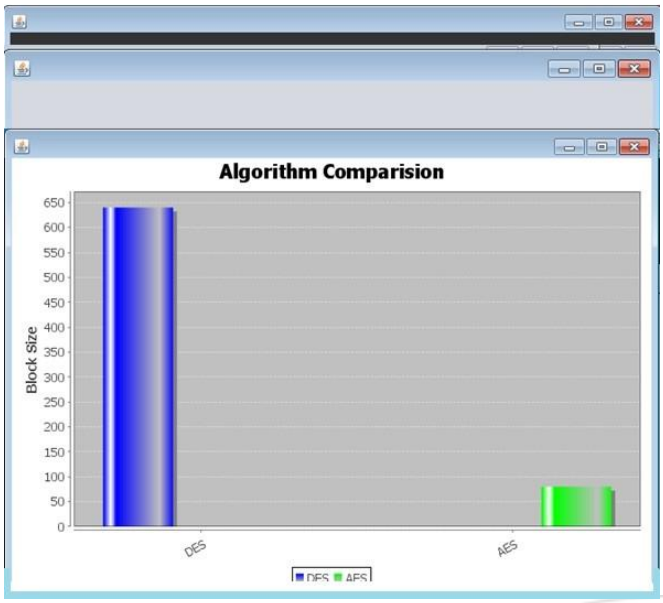

Figure 4: Result Analysis

7. Conclusion

In this paper, we propose secure hashing algorithm for avoiding deduplication, which generates a unique key for each file. If a slight change is occurs in new file, whole key will be changed. Hence, strict dedeplication can be possible. Also we use encryption techniques for providing security inside the public cloud. Hence, we can protect data from unauthorized access.

8. Future Scope

We plan to investigate the secure deduplication issue in cloud backup services of the personal computing environment. We can further explore and exploit index lookup parallelism availed by the application-aware index structure of Deduplication in multi core environment.

9. Acknowledgment

I wish to express my sincere thanks and deep gratitude towards Dr. Y.V. Chavan [Principal PVPIT, Pune] and my guide Mr. Navnath D. Kale for his guidance, valuable suggestions and constant encouragement in all phases. I am highly indebted to his help in solving my difficulties which came across whole Paper work. Finally I extend my sincere thanks to respected Head of the department Mr. N. D. Kale and all the staff members for their kind support and encouragement for this paper. I extend my thanks to cPGCON co-ordinators and organizers. Last but not the least, I wish to thank my family for their unconditional love and support.

References

[1] Jin Li and Yan Kit Li, A Hybrid Cloud Approach for Secure Authorized Deduplication, March 2014.
[2] P. Anderson and L. Zhang, Fast and secure laptop backups with encrypted deduplication, In Proc. of USENIX LISA, 2010.
[3] M. Bellare, S. Keelveedhi, and T. Ristenpart, Dupless: Server aided encryption for deduplicated storage, In USENIX Security Symposium, 2013.
[4] M. Bellare, S. Keelveedhi, and T. Ristenpart, Message-locked encryption and secure deduplication, In EUROCRYPT, pages 296312, 2013.
[5] M. Bellare, C. Namprempre, and G. Neven, Security proofs for identity- based identification and signature schemesn, J. Cryptology,22(1):161, 2009.
[6] M. Bellare and A. Palacio, Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks, In CRYPTO, pages 162177, 2002.