www.ijecs.in

International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 3 March 2017, Page No. 20602-20605
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i3.36

Security Based Deduplication with Efficient and Reliable Data Storage

J.Sreejith, A. Jeno Kalaiselvi Sathya
(MCA, PSN College &Technology, Tirunelveli, Tamilnadu-627152, India)
Email ID: sreejithjayachandran012@gmail.com
(Assistant professor, PSN College &Technology, Tirunelveli, Tamilnadu-627152, India)
Email ID*: jenored@gmail.com*

Abstract - Data is a set of values of quantitative variable information. Data deduplication means that the data compression technique for eliminating duplicate copies of repeating data. In the deduplication process, unique types of data, or byte patterns or images, which are identified and stored during a process of analysis Data deduplication is increasingly accepted as a technique to reduce storage costs. In cloud computing system the data deduplication acts as a special role. The Cloud computing is a type of Internet-based computing, that provides shared computer processing that means data to computers and other devices on demand. The example of cloud services are computer networks, servers, storage, applications and in these most commonly using cloud service is data storage. The data are often stored in cloud in an encrypted form. One application of cryptographic splitting is to provide security for cloud computing. However encrypted data introduce new problems arise in cloud data deduplication, which is big data storage and processing. Existing system suffer from security weakness. They cannot flexibly support data access control and revocation. The cloud-based storage example such as Drop box, Google Drive, and Mozy can save on storage costs via deduplication. We make a data will be secure using a user key before he store in to the cloud. But common encryption modes are randomized. People upload personal or confidential data to the data center of a Cloud Service Provider (CSP) but it is do not give fully secure for the data because the data holders may not be always online or available for such a management, which could cause storage delay. And also create the Dekey is secure in terms of the definitions specified in the proposed security model.

Keywords:-cloud service provider, Data storage, Deduplication, cryptography, DeKey.

1. INTRODUCTION

Cloud computing means that network of remote servers hosted on the Internet to store, manage, and processing in the data, mainly many of providers of cloud-based storage such as Drop box, Google Drive and Mozy these are helps to save on storage costs via deduplication: it means that the two clients upload the same file; the service detects that data or files and stores only a single copy. This processes mainly helping to the cloud servers. The cloud computing contains scalability, elasticity, fault-tolerance, and pay-per-use. The cloud computing mainly cryptographic splitting is to provide for security. A client could encrypt its file, under a user's key, before storing it. But common encryption modes are randomized; making deduplication impossible since the SS (Storage Service) effectively always sees different cipher texts regardless of the data. The advent of cloud storage motivates enterprises and organizations to outsource data storage to third party cloud providers. One of the critical challenges of today's cloud storage services is the management of the ever-increasing volume of data. In 2020 the volume of data will be reach in approximately 40 trillion gigabyte it is huge. Deduplication is a technique to reduce the storage space and increasing the bandwidth in cloud storage so they make data management will be scalable one. Instead of keeping multiple data copies with the same content deduplication eliminates redundant data by keeping only one physical copy. it has been checking the whole file it is called file level deduplication, or a more fine-grained fixed size or variable-size of data block that is called block-level deduplication .Today's commercial cloud storage services, such as Drop box, Mozy, and Memo pal, have been applying deduplication to user data to save maintenance cost and also in storage problem .From a user's perspective, data outsourcing raises security and privacy concerns. In third-party cloud providers to properly enforce confidentiality, integrity checking, and access control mechanisms against any insider and outsider attacks. From a user's point of view, data outsourcing raises security and privacy concerns. Deduplication process is applying while improving storage and bandwidth efficiency is compatible with Convergent key management it's using for security reason. Specifically, traditional encryption requires different users to encrypt their data with their own keys. And mostly the hackers do not attack while we do the important file in different location in cloud. The basic idea of security is decrease the value of the stolen information to the attacker. And achieve this through a 'preventive' disinformation attack.

1.1 File-level deduplication: file-level deduplication means that the multiple copies of the files, stores that already saved and then just links the other references to the first file. And

DOI: 10.18535/ijecs/v6i3.36

we can get only one copy of file related. This is file-level duplication.

1.2 Block-level deduplication: in block level deduplication means the data that we want to save that already exists the block. If so, the second copies are not stored on the disk/tape, but it created to point to the original copy, this is known as block-level deduplication.

2. RELATED WORKS

2.1 Finding duplicate data in deduplication

They are two types of data deduplication process one is block level and other is file level deduplication now a day's widely deduplication is using for minimizing disk capacity and reduce network traffic. They are two names Multi-mode Venti is a network storage system with a 160-bit SHA-1 hash of the data that enforces a write-once policy since no other data block can be found with the same address. So duplicate data is easily identified and the data block is stored only once. LBFS (Low Bandwidth File System) it avoids already send the data over the network when the same data can already be found in the server's file system or the client's cache and increasing the bandwidth of the system.

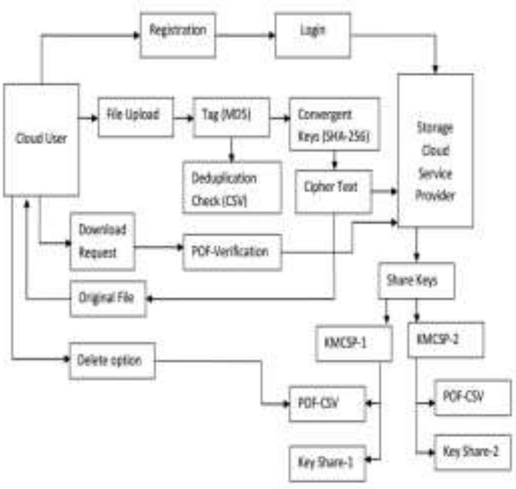
2.2 Using Convergent Encryption for security

Encryption using for security and here using the convergent encryption that is derived from the hash of the plain text and in simple way of convergent encryption example is; sanal derives the encryption key from his message F such that D = A(N), where A is a cryptographic hash function; she can encrypt the message with this key, hence: R = C(D, N) = C(A(N), N), where C is a block cipher. This technique helping for detecting, two users with two identical plaintexts will obtain two identical cipher texts since the encryption key is the same; then cloud storage provider will be able to perform deduplication on such cipher texts. And the encryption keys are generated, retained and protected by users. As the encryption key is deterministically generated from the plaintext, users do not have to interact with each other for the key to encrypt for a given plaintext. Therefore, convergent encryption is better for candidate, for the adoption of encryption and deduplication in the cloud storage domain. The security weaknesses of Convergent Encryption is mainly, the requirement for deduplication at block-level further raises an issue with respect to key management. As an inherent feature of CE, the fact that encryption keys are derived from the data itself does not eliminate the need for the user to memorize the value of the key for each encrypted data segment in file-level deduplication. In block-level deduplication the requirement to memorize and retrieve CE keys for each block in a secure way, calls for a fully-fledged key management solution and also including a new component, the metadata manager (MM), in the new ClouDedup system in order to implement the key management for each block together with the actual deduplication operation for better performance.

2.3 Storage into cloud

Storage manager stores tag of file, that already creating for security reason that tag is not already present into the cloud. Then tags formation function is simply called chunks. So the chunk is redundant then link to the chunk is returned to the client by updating metadata for the chunk or file. We can see as in index, it is stored according to the type of application. This will be reduced the size of the files and the disk lookup overhead for deduplication by dividing big index to application level indexes. This the way to storage into cloud after deduplication. Cloud storage plays an important role in cloud computing.

3. PROPOSED WORKS



Dekey a new structure in which users do not need to administer any keys on their own but as an alternative securely distribute the convergent key shares across compound servers. Dekey the term of deduplication effectively and security and it is also the Proof of Ownership of the file. The outsource the convergent keys to third party key Management server securely. DeKey supports both file-level and block level using the secret sharing scheme and make obvious that Dekey incurs incomplete overhead in realistic upbringing so we create a new construction called Dekey which make available efficiency and dependability undertaking for convergent key organization on both user and cloud storage. Data deduplication is a specialized data compression technique for using eliminating duplicate data. This technique is help to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent in cloud. In this process, the unique data, or byte patterns, are identified and stored during a process of analysis after the analysis continues , it is compare to the other chunks that already stored whenever a match occurs, the redundant chunk is replaced with a small reference that points to the stored chunk and given that the

DOI: 10.18535/ijecs/v6i3.36

same byte pattern may occur dozens or hundreds, or even thousands of times (the frequency is calculate dependent on the chunk size), the amount of data that must be stored or transferred can be reduced it is also secure and another is convergent encryption. The convergent encryption (CE) means the encryption key from the hash of the plaintext. The convergent encryption is a technique, that apply in between the two users with two similar plain text will obtain two similar cipher texts since the encryption key is the same; then the cloud storage provider will be able to perform deduplication on such cipher texts .The encryption keys are generated, retained and protected by users. New component, the metadata manager (MM), in the new ClouDedup system is implement the key management for each block according to their order and together with the actual deduplication operation and the role of the user is limited to splitting files into blocks, encrypting them with the convergent encryption technique, signing the resulting encrypted blocks and creating the storage request. In addition, the user also encrypts each key derived from the corresponding block with the previous one and his secret key in order to outsource the keying material as well and thus only store the key derived from the first block and the file identifier. For each file, this key will be used to decrypt and re-build the file when it will be retrieved. Instead, the file identifier is necessary to univocally identify a file over the whole system.

The storage phase in the user computes the signature of the hash of the first block is

$$S0 = \sigma PKu\ (H\ (B0)) \quad (1)$$

The monitor data access in the cloud and detect abnormal data access patterns. Metadata Manager (MM) is the component responsible for storing metadata, which include encrypted keys and block signatures, and handling deduplication.MM maintains a linked list and a small database in order to keep track of file ownerships file composition and avoid the storage of multiple copies of the same data segments. The tables used for this purpose are file, pointer and signature tables. Each node in the linked list represents a data block. The identifier of each node is obtained by hashing the encrypted data block received from the server. If there is a link between two nodes X and Y, it means that X is the predecessor of Y in a given file. A link between two nodes X and Y corresponds to the file identifier and the encryption of the key to decrypt the data block Y. They are three types 1) File table,2)pointer table,3)signature table. In file table contains the file id, file name, user id and the id of the first data block. The pointer table contains the block id and the id of the block stored at the cloud storage provider and the signature table contains the block id, the file id and the signature.

4. CONCLUSION & FUTURE ENHANCEMENT

The basic idea of secure deduplication services can be implemented given additional security features insider attacker on Deduplication and outsider attacker by using the detection of masquerade activity which means unknown person stolen and damage the data. So we confusion of the attacker and the additional costs incurred to distinguish real from fake information added, and the deterrence effect which, although hard to measure, plays a significant role in preventing from the attackers , that will harmful for our data .The combination of these security features will provide extreme levels of security for the deduplication. So we create DeKey, an efficient and reliable convergent key management scheme for secure deduplication. DeKey applies deduplication among convergent keys and distributes convergent key and shares multiple keys, the convergent keys are using for security and confidentiality of outsourced data. Mainly the DeKey is using for Proof of ownership scheme and demonstrate that it incurs small encoding/decoding overhead compared to the network transmission. The DeKey helps regular upload/download operations too. In future enhancement using better dedupe algorithms, cloud and tape integration, improved accuracy, improved scalability, global deduplications. In future they have adigitilaized world so all the data save into virtual means cloud so the size of cloud storage will be increasing propositional to data increasing.

6. REFERENCES

[1] D. Perttula, B. Warner, and Z. Wilcox-O'Hearn, "Attacks on convergent encryption." (2016). [Online]. Available: http://bit.ly/ yQxyvl

[2] M. Bellare, S. Keelveedhi, and T. Ristenpart, "DupLESS: Server aided encryption for deduplicated storage," in Proc. 22nd USENIX Conf. Secur., 2013, pp. 179–194.

[3] Google Drive. (2016). [Online]. Available: http://drive.google.com

[4] Mozy, Mozy: A File-storage and Sharing Service. (2016). [Online]. Available: http://mozy.com/

[5] C. Yang, J. Ren, and J. F. Ma,"Provable ownership of file in deduplication cloud storage," in Proc. IEEE Global Commun. Conf.,2013, pp. 695–700, doi:10.1109/GLOCOM.2013.6831153.

[6] P. Puzio, R. Molva, M. Onen, and S. Loureiro, “ClouDedup:Secure deduplication with encrypted data for cloud storage,” in Proc. IEEE Int. Cof. Cloud Comput. Technol. Sci., 2013, pp. 363–370,doi:10.1109/CloudCom.2013.54
.
[7] Dropbox, A file-storage and sharing service. (2016). [Online].Available: http://www.dropbox.com.

AUTHOR PROFILE

J.Sreejith received the B.SC. Degree in computer science from SreeSankara Vidyapeedam College Nagaroor,Kerala university in 2014, respectively. During 1997-1999, now I’am doing MCA in, PSN College &Technology, Tirunelveli, Tamilnadu-627152.

A. Jeno Kalaiselvi Sathya received the B.Sc Degree in Rani Anna Government College for women, Tirunelveli in 2001 and MCA from IndhraGandhi National Open University in 2005& ME Computer Science in PSN College &Technology, Tirunelveli,in 2010.