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

IJCSMC, Vol. 4, Issue. 1, January 2015, pg.421 – 425

SURVEY ARTICLE

Survey on Authorized Data Deduplication System using Cryptographic and Access Control Techniques

Santoshi S Patil[1], Samprati T[2], Asst. Prof. Shwetha K S[3]
Student, M.Tech (Software Engineering), NHCE, Bangalore, India[1]
Student, M.Tech (computer science), AIT, Bangalore, India[2]
Jr. Assistant Professor, Information Science & Engineering Department, NHCE, Bangalore, India[3]
Santu91patil@gmail.com; sampratituppad@gmail.com

Abstract- Ever increasing volume of back up data in cloud storage may be a vital challenge .back up windows are shinking due to growth of information .we use the concept of deduplicate. Deduplication means duplicate data is eliminated a pointer is created to reference a data that is backed up. Deduplication can take place at file level, in this it detects redundant data within and across files or at the block level, in this it removes redundant copies of identical files.

Keywords: Deduplication, secure storage, cryptographic

I. Introduction

The increasing volume of data has raised a vital problem for data protection in the personal computing environment .thus cloud computing concept has been used for big storage of data and sharing of resources over a network. Where we have a large pools of resources can be connected .this affiliation may be private or public. A private cloud is a virtualized data center that operates within the firewall .in public services are available generally used over the network

Cloud storage is getting popular more and more as it is low cost and on-demand use of large storage .to make data management scalable in cloud computing, we use the deduplication concept. Data deduplication concept that helps to reduce the cost associated with large scale data backups.

For example, if messages on associated degree organization's email server are being backed up and a mass email with an attachment has been sent to 50 employees of the organization, only one copy of the email will be backed up, since the other 49 instances are duplicates.

In order to have a secure storage of deduplicated data over a cloud computing we use the encryption/decryption technique. Encryption is the process of converting a plaintext into a ciphertext. Decryption is the process of converting a ciphertext into plain text. The process of symmetric key encryption where same key is used for both encryption and decryption .the key ought to be send firmly over a network. To stop unauthorized access, a secure proof of possession protocol is additionally required to provide the proof that the user indeed owns the similar file when a duplicate is found. Once the proof, consequent users with the similar file are going to be provided a pointer from the server while not having to transfer the similar file.

II. Security Issues with Deduplication Technique

Encryption is the process of converting a plaintext into a ciphertext. Decryption is the process of converting a ciphertext into plain text. Since user encrypt their files with their individual encryption key, different ciphertext would emerge, even for identical files .thus encryption fails for deduplication concept.

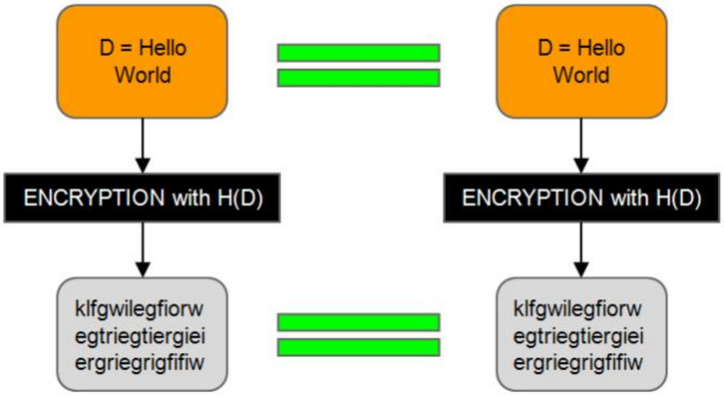
D = Hello World
D = Hello World
ENCRYPTION with H(D)
ENCRYPTION with H(D)
klfgwilegfiorw egtriegtiergiei ergriegrigfifiw
klfgwilegfiorw egtriegtiergiei ergriegrigfifiw

Figure.1 Convergent encryption

Convergent Encryption concept has been used. In convergent encryption, the encryption key is derived from the data .This way, two different users will automatically use the same encryption key for the same data and thus generate the same ciphertext.as shown in figure1. This means that we have found the solution to combine deduplication and encryption.
But the disadvantage over this convergent encryption is the key of a given data segment can be generated by anyone in a deterministic way. An attacker can generate a key from a plaintext, encrypt the plaintext and check if the resulting ciphertext is already stored or not.

Convergent encryption has been overcome by adding one additional layer of deterministic and symmetric encryption on top of convergent encryption .This additional encryption can be added by a component placed between the user and the cloud storage provider such as a local server or a gateway. This component will provide encrypting and decrypting of data from and to users. In order to allow the cloud provider to detect duplicates, encryption and decryption operation is done with a unique set of secret key.

We suggest to introduce a new component which we called **metadata manager**. The goal of this additional component is to store encrypted block keys and perform deduplication on encrypted blocks. Thanks to this separation between data and metadata. we achieve the **complete independence from the storage provider.** A metadata manager that updates the metadata (in order to rebuild the structure of each file) , stores encrypted block keys and performs deduplication on encrypted blocks. Only those blocks that are not already stored are actually stored.

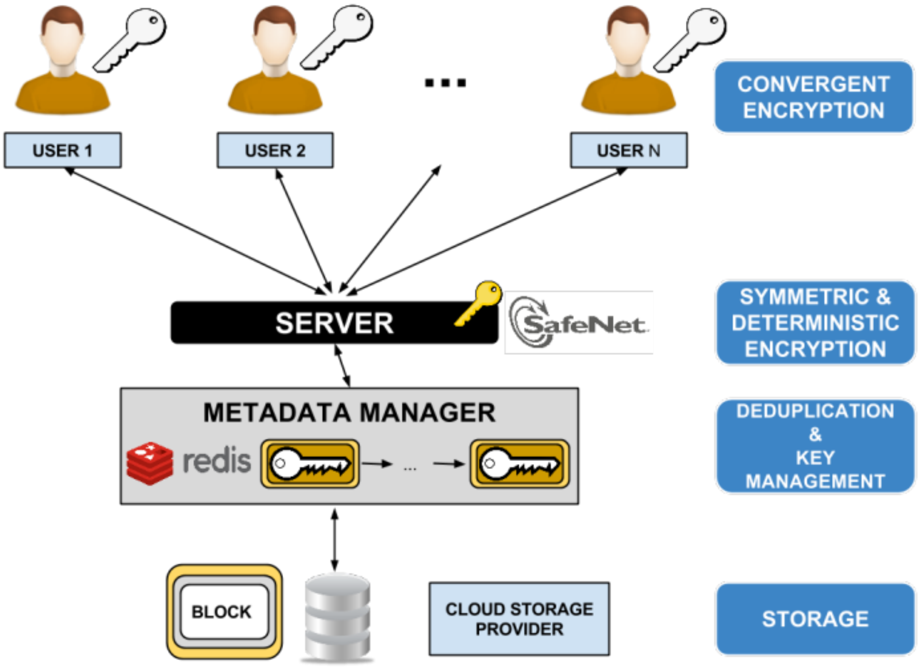

Figure.2 Additional layer of deterministic and symmetric encryption on top of convergent encryption

III. Techniques for providing security to deduplicat data storage.

In this paper[1] they have explained the convergent encryption, the data copy is encrypted under a key derived by hashing the data itself. This convergent key is used for encrypting and decrypting of a data copy. After key generation and data encryption. The used the deterministic encryption process where identical data copies will generate the same convergent key and the same cipher text. Thus convergent encryption allows to perform deduplication on cipher texts. The cipher texts can be decrypted by the corresponding data owners only with their convergent keys. Duplicate check authority is given only for a authorized user .a set of privileges is given based on their authorization during system initialization. This set of privileges specifies that which kind of users is allowed to perform duplicate check and access the files.

In this paper[2] they proposed the policy based deduplication scheme.in order to limit the ability and the knowledge needed for deduplication and To have a certain degree of security assurance they proposed a policy-based de-duplication proxy scheme to enable different trust relations among cloud storage components. The de-duplication proxy scheme will de-duplicate data of its registered users based on the capabilities it has received from the key center. The user can decide which data will be de-duplicated by submitting the tags of the data chunks

In this paper[3] they proposed An Application-Aware source deduplication[AA] as the process of cloud back up system is increasing ,deduplication can achieve high space efficiency and it reduces the cloud storage cost. The process of cloud efficiency becomes critical for cloud clients in the personal computing environment due to its limited system resources. To achieve high space efficiency EMC Avamar[24] applies CDC based chunk level deduplication with high computational overhead and lookup overhead .In this AA-deduplication improves deduplication efficiency significantly by intelligent data chunking methods with application awareness. But exploits application awareness by limiting the search for redundant data to the chunks within the same kind of applications specified by the file format information. The direction for future work ,they planned to investigate the secure deduplication issues in cloud backup services of personal computing environment and further explore and exploit index lookup parallelism by application aware index structure of AA deduple in a multi-core.

In this paper[4] they have explained how to provide security even for lock dependent messages. first approach is to avoid using tags that are derived deterministically from the message .to this end ,we design a fully randomized scheme that supports an equality-testing algorithm defined on the cipher texts. They design an algorithm that encrypt message under a key that is highly correlated with message and still remain secure. secondly the part of ciphertext that allows the equality test must not leak any information about message from an adversarially chosen min-entropy distribution even given the public parameters.

In this paper[5] they explained how to achieve storage efficiency, limited memory usage for deduplication indexing, and to achieve high throughput of multiple backup streams using RevDedup .revdedup which removes duplicates of old backups and mitigates fragmentation of latest backups.RevDedup applies global deduplication over a large size data units.to maintain high deduplication efficiency, it maintain the data placement as sequential as possible for the latest version, and removing any redundant data of old versions and referring it to the identical data Of the latest version.

In this paper [6]They have explained about the number of security issues in cloud computing as it encompasses many technologies including network ,databases, operating system, resource scheduling, transaction management, load balancing. For example, the network that interconnecting the systems in a cloud has to be secured and mapping the virtual machine to the physical machines has to be carried out securely. Cloud service providers need to inform their customers the level of security that they provide on their cloud.

In this paper[7] they have explained about the how to design and implement a fast and secure data backup process. They used the technique of one time password ,it is only perfect encryption .the sender and receiver must each have a copy of same pad ,that has to be transmitted over a secure line. The pad used as a symmetric key .once pad is used, it is destroyed it makes perfect for extremely high security situation ,but this technique is unusable for everyday use.

IV. Conclusion and future work

In this paper we have analyzed concerning deduplication concept, where we have a tendency to conferred many new duplicate check in cloud architecture and access privileges given to retrieve or store the data in cloud. Some of the security issues, where security fails during deduplication thus they proposed a convergent encryption scheme to provide a secure deduplication storage process. We have analyzed through different tools for data security i.e. convergent encryption scheme, symmetric encryption scheme, one time password scheme.

Future work is to have a authorized deduplication and we provide a rank to authorized user based on their privilege .to have secure storage we use the concept of hybrid cloud concept where we have a private cloud and public cloud .this Reduces the load of a cloud server in a huge amount as duplicate files are not uploaded. Private cloud entity introduce for facilitating user’s se-cure usage of cloud service. The speed of the server is increased.

References

[1] Reclaiming Space from Duplicate Files in a Serverless Distributed File System. John R.Douceur, Atul Adya, William J. Bolosky, Dan Simon, Marvin Theimer Microsoft .

[2] A Policy-based De-duplication Mechanism for Encrypted Cloud Storage Chuanyi LIU 1, Yancheng WANG2, Jie LIN.

[3] Application-Aware Client-Side Data Reduction and Encryption of Personal Data in Cloud Backup Services. Yin-Jin Fu1 , Nong Xiao, Xiang-Ke Liao, *Member*, Fang Liu.

[4] Message-Locked Encryption for Lock-Dependent Messages. Mart.n Abadi, Dan Boneh, Ilya Mironov,Ananth Raghunathan, and Gil Segev2.

[5] RevDedup: A Reverse Deduplication Storage System Optimized for Reads to Latest Backups.Chun-Ho Ng and Patrick P. C. Lee.

[6] Cloud Computing: Security Issues and Research Challenges. Rabi Prasad Padhy, Manas Ranjan ,PatraSuresh,Chandra Satapathy.

[7] Optimal Authorized Data Deduplication in Cloud.Divyesh Minjrola , Rakesh Rajani .

[8] Server-Aided Encryption for Deduplicated Storage. Mihir Bellare *University* ,Sriram Keelveedhi, *San Diego*,Thomas Ristenpart.

[9] Distributed Key Generation for Secure Encrypted Deduplication.Yitao Duan,NetEase Youdao.

[10] A fusion Cloud approach for Certified Deduplication. S.Md.Samiullah1, B.Jagadeesh2 , S.Md.Hafeez3 & D.Jayanarayana Reddy4 .

[11] Secure Data Deduplication. Mark W. Storer Kevin Greenan Darrell D. E. Long Ethan L. Miller.

[12] Proceedings of the 5th Symposium on Operating Systems Design and Implementation. Boston, Massachusetts, USA December 9–11, 2002

[13] A Secure Cloud Backup System with Assured Deletion and Version Control. Arthur Rahumed, Henry C. H. Chen, Yang Tang, Patrick P. C. Lee, and John C. S. Lui. The Chinese University of Hong Kong, Hong Kong.

[14] Secure Audit Service by Using TPA for Data Integrity in Cloud System. Shingare, Vidya Marshal.