IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 2, February 2016

Protecting Outsourced Data in Cloud Using Deduplication and Enhancing Key Management

Anuja Kunjumon[1], Devu S Nair[1], Saranya P[1], Shahanas Abdul Sammad[1], Smitha C S[2]

U.G. Student, College of Engineering Perumon, Kerala, India[1]

Assistant Professor, Dept of Computer Science and Engineering, College of Engineering Perumon, Kerala, India[2]

Abstract: Data Deduplication is a technique used in cloud storage to remove duplicate copies of similar data. The main motive of deduplication is to reduce storage space and improve bandwidth. In order to provide security and privacy to outsourced data encryption techniques are used. Encryption is performed by generating keys. To attain efficiency and reliability in managing a huge number of keys, a scheme called Dekey is introduced. Dekey is implemented using ramp Secret Sharing Scheme. Dekey provides a way to reduce overhead by distributing convergent keys across multiple servers. This paper also proposes a means of sharing data among multiple users.

Keywords: Data Deduplication, Encryption, Dekey, Ramp Secret Sharing Scheme.

I. INTRODUCTION

The emergence of cloud storage has motivated many enterprises to outsourced data storage to many third party cloud service providers as done by Amazon, Google, and IBM. Even though many are demanding for cloud storage, there are a number of challenges faced by it. Increasing data volume, security and privacy are some of the major concerns. To reduce the amount of data stored, a technique called Deduplication has been used. It improves the bandwidth and saves maintenance cost along with scalable data management. In this technique, a single copy of the data content is stored rather than storing multiple copies of the same data and a reference is given to all the redundant data.

Deduplication can be of File level deduplication or Block level deduplication. In File level deduplication redundant files are eliminated whereas in Block level deduplication each file is fragmented and the redundant fragments are removed. Outsourcing data to the cloud faces many security and privacy issues. Data Integrity, Confidentiality, Data theft, Data loss is some of these issues. In order to ensure confidentiality, integrity and access control mechanism, a trusted third party cloud providers are essential.

To provide security in cloud, various encryption techniques can be used. In traditional encryption, data is encrypted using users own key. Using Traditional encryption [8] [9] [10] deduplication is not possible as encryption using different keys produce different ciphertexts Convergent encryption [1] is another technique, where data encryption as well as decryption takes place using the same key derived by computing the cryptographic hash value of the data content itself. Convergent key is encrypted using user's master key. Encrypted convergent key is stored in the cloud along with the data and the master key is maintained by the users.

This method is highly unreliable and ineffective as the number of user's increases; the number of keys also increases. Also if the master key is lost then it becomes impossible to gain access to outsourced data stored. In order to overcome unreliability and inefficiency a new technique called Dekey is introduced. Dekey is used to efficiently manage the convergent key.

II. LITERATURE REVIEW

By examining different primitives needed to perform secure deduplication, Encryption, Proof of Ownership, Ramp secret Sharing Scheme.

A. Data Deduplication

Data deduplication [5] [6] [7] is a technique adopted to eliminate multiple copies of similar data by maintaining a single copy and giving a reference to one physical data. This technique is commonly used in cloud storage to reduce storage space and to improve bandwidth. Different organisations use this technique to maintain system backups. Deduplication can be File level and Block level deduplication. File level deduplication eliminates identical copies of same file whereas in block level, a file is divided into different blocks and eliminates identical blocks.

B. Encryption

To provide security and privacy to the outsourced data encryption techniques are used before outsourcing the data.

i) Traditional Encryption:

Traditional encryption encrypts/decrypts the data using different keys. So the resulting cipher text will be different. So deduplication cannot be performed in traditional encryption.

ii) Convergent Encryption:

Convergent encryption [1] [11] encrypts /decrypts the data using a convergent key that is generated from the data itself. The identical data will generate same convergent key and resulting cipher text will also be same. So deduplication can be performed easily using convergent encryption.

DOI 10.17148/IJARCCE.2016.52126

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 2, February 2016

C. Proof Of Ownership

The notion of proof of ownership (POW) [4] will help the users to prove their identity of ownership of a data to storage cloud service provider. This mechanism protects the security in a client side deduplication. Generally the proof of ownership is implemented as an iterative algorithm run by the user entity and storage cloud service provider.

D. Ramp Secret Sharing Scheme

Ramp Secret Sharing Scheme (RSSS) [3] is used to store the convergent keys in Dekey. (n, k, r)-RSS scheme will generate n shares from a secret such that any k shares can recover the secret but from any r shares the information about the secret cannot be deduced. When r=0 (n, k, 0)-RSS become Rabin's Information Dispersal Algorithm (IDA) and when r=k-1, the (n, k, k-1)-RSS becomes (n, k)-Shamir Secret Sharing Scheme (SSSS).

III.SYSTEM IMPLEMENTATION

We formulate a system model that provides security and privacy to the outsourced data through encryption and deduplication techniques. Dekey is used to efficiently manage the convergent keys. The Fig.1 represents the flow diagram of the generation of Dekey.

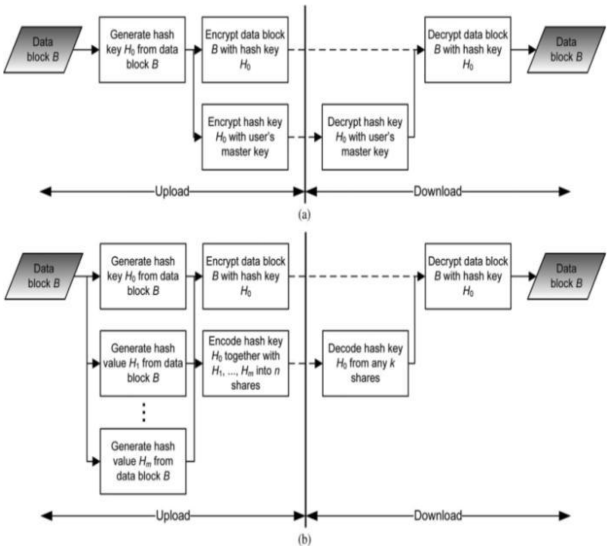
Data block B
Generate hash key H_0 from data block B
Encrypt data block B with hash key H_0
Encrypt hash key H_0 with user's master key
Decrypt hash key H_0 with user's master key
Decrypt data block B with hash key H_0
Data block B
Upload
Download
(a)
Data block B
Generate hash key H_0 from data block B
Generate hash value H_1 from data block B
Generate hash value H_m from data block B
Encrypt data block B with hash key H_0
Encode hash key H_0 together with $H_1, ..., H_m$ into n shares
Decode hash key H_0 from any k shares
Decrypt data block B with hash key H_0
Data block B
Upload
Download
(b)

Fig1. Flow diagram of the generation of Dekey

There are three main entities namely User, Storage Cloud Service Provider(S-CSP) and Key Management Cloud Service Provider (KM-CSP).

User: This entity outsources the data to a storage cloud service provider and access the data later.

S-CSP: This entity provides data outsourcing facilities and store data on behalf of the user. In order to reduce storage space and upload bandwidth, S-CSP eliminates redundant copies of data via deduplication by maintaining a single copy and giving reference to all redundant data.

KM-CSP: This entity maintains convergent keys for users. Each convergent key is distributed across multiple KM-CSP using Ramp Secret Sharing Scheme.

A. Upload

In order to outsource a data, first we need to encrypt it using the key generated from the data content itself. Initially a tag is generated for the entire data and proceeded by generating sub tags for each of the blocks. A unique tag id is generated for unique data. The generated tags are then provided to the S-CSP who then examines whether the same tag exist. If yes, then file level or block level deduplication takes place.

According to Dekey concept,

- If two or more user uploads different files, then data is encrypted using different convergent keys.
- If two or more users upload similar files, then it generates some convergent keys and it is shared between users with similar files. Then the convergent key is divided into shares and distributed among multiple KM-CSP.
- If two or more users upload different files with one or more than one similar sub tags, then it generates different convergent keys. In order to encrypt/decrypt files containing similar sub tags, convergent key must be shared.

B. Download

In order to download a file, the user first requests the S-CSP for a file. If the user is authenticated, the S-CSP decrypts the encrypted blocks by recovering the convergent keys from multiple KM-CSP. Then the decrypted blocks of data are reconstructed to obtain the original files.

C. Sharing

If a user wants to access a file belonging to different user, the user sends a request to the owner of the file. This request is first verified by the S-CSP by means of authentication. S-CSP then forwarded the request to the owner. If the owner accepts the request, then permission is granted for accessing the file else rejected.

IV. RESULT ANALYSIS

The file uploading done with the size in the range of kilo bytes hashing for a particular blocks. The hashing is improved with a higher file size in comparatively minimum response time. The security aspect can be improved by constructing efficient deduplication and key aspects. The data sharing is improved with high security enhancements.

V. CONCLUSION

The proposed model provides an efficient and reliable key management scheme in order to perform secure deduplication. It uploads bandwidth and reduces storage space. Key management is performed using Dekey technique. Dekey is implemented using Ramp Secret Sharing Scheme. It provides privacy and security to the outsourced data.

DOI 10.17148/IJARCCE.2016.52126

REFERENCES

[1] J.R. Douceur, A. Adya, W.J. Bolosky, D. Simon, and M. Theimer, ‘‘Reclaiming Space from Duplicate Files in a Server less Distributed
[2] File System,’’ in Proc. ICDCS, 2002, pp. 617-624.
[3] J. Breckling, Ed., the Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.
[4] G.R. Blakley and C. Meadows, ‘‘Security of Ramp Schemes,’’ in Proc. Adv. CRYPTO, vol. 196, Lecture Notes in Computer Science,G.R. Blakley and D. Chaum, Eds., 1985, pp. 242-268
[5] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg,
[6] ‘‘Proofs of Ownership in Remote Storage Systems,’’ in Proc. ACM Conf. Comput. Commun. Security, Y. Chen, G. Danezis, and V. Shmatikov, Eds., 2011, pp. 491-500
[7] P. Anderson and L. Zhang, ‘‘Fast and Secure Laptop Backups with Encrypted De-Duplication,’’ in Proc. USENIX LISA, 2010, pp. 1-8.
[8] 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 Proc. 3rd Int’l Workshop Security Cloud Comput., 2011, pp. 160-167
[9] M.W. Storer, K. Greenan, D.D.E. Long, and E.L. Miller, ‘‘Secure Data Deduplication,’’ in Proc. StorageSS, 2008, pp. 1-10.
[10] W. Wang, Z. Li, R. Owens, and B. Bhargava, ‘‘Secure and
[11] Efficient Access to Outsourced Data,’’ in Proc. ACM CCSW, Nov. 2009, pp. 55-66.
[12] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, ‘‘Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing,’’ IEEE Trans. Parallel Distrib. Syst., vol. 22, no. 5, pp. 847-859, May 2011.
[13] Yun, C. Shi, and Y. Kim, ‘‘On Protecting Integrity and
[14] Confidentiality of Cryptographic File System for Outsourced Storage,’’ in Proc. ACM CCSW, Nov. 2009, pp. 67-76. Jin.
[15] M. Bellare, S. Keelveedhi, and T. Ristenpart, ‘‘Message-Locked Encryption and Secure Deduplication,’’ in Proc. IACR Cryptology ePrint Archive, 2012, pp. 296-3122012:631.

DOI 10.17148/IJARCCE.2016.52126