ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Secure Data Storage with Deduplication and Efficient Convergent Key Management

Soofiya MS[1], Sreetha V Kumar[2]

M. Tech Student, Marian Engineering College, Trivandrum, Kerala, India[1]

Asst. Professor, Dept. of CSE, Marian Engineering College, Trivandrum, Kerala, India[2]

ABSTRACT: Businesses and consumers are becoming increasingly conscious of the value of secure archival data storage. Most of the organization need to outsource their confidential data including sensitive information to cloud. One of the greatest challenge todays cloud storage services facing is the management of ever increasing volume of data and most importantly security. This paper introduces a technique called deduplication which avoids duplicates copies of data been stored in the cloud thus saving upload bandwidth and space for storing data. One of the extensively adopted technique for secure deduplication is the convergent encryption method. The main issue in making convergent encryption practical is the management of huge number of keys with increase in number of users. The basic idea of this paper is to secure data storage and avoid redundant copies. To this end we propose a new technique called Dekey using visual cryptography, DCT image compression for securing and reducing data being stored. Dekey is a technique in which user do not want to manage the keys instead keys are distributed in multiple servers.

KEYWORDS: deduplication, convergent encryption, visual cryptography, DCT image compression

I. INTRODUCTION

Data deduplication is a well-known technique for management of data by eliminating the redundant copies being stored. In deduplication only one physical copy is kept and other copies are referred to this physical copy. Although convergent encryption is used for deduplication the main challenge is the efficient management of keys with increase in number of users.

Each party that outsources their data to cloud encrypt the data with different keys for sake of confidentiality producing different cipher text even for same data thus making deduplication impossible [7]. Convergent encryption provides an option for data confidentiality and reliability while making Deduplication possible.For implementing secure storage additional to deduplication the data is embedded in an image. The image is then compressed using the DCT image compression technique [4].

This paper makes the following contribution:

A new construction Dekey [1] is proposed for efficient and reliable convergent key management through convergent key deduplication and secret sharing.

Security analysis demonstrates Dekey is secure in terms of definition specified in proposed security model.

We implement Dekey using visual cryptographic method that enables the key management to adopt to different confidentiality levels.

DCT image compression is used to reduce the upload bandwidth.

II. RELATED WORK

As the world moves to digital storage for archival purposes, there is an increasing demand for systems that can provide secure data storage in a cost-effective manner. By identifying common chunks of data both within and between files and storing them only once, deduplication can yield cost savings by increasing the utility of a given amount of storage. Unfortunately, deduplication exploits identical content, while encryption attempts to make all content appear random; the same content encrypted with two different keys results in very different ciphertext. Thus, combining the space efficiency of deduplication with the secrecy aspects of encryption isproblematic [9].To protect the confidentiality of outsourced data, various cryptographic solutions have been proposed in the literature Their idea builds on traditional

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

(symmetric) encryption, in which each user encrypts data with an independent secret key. Some studies propose to use threshold secret sharing to maintain the robustness of key management. However, the above studies do not consider deduplication. In traditional encryption different users use their own secret keys for encryption resulting in different cipher texts thus making deduplication impossible [2]. Convergent encryption ensures data privacy and enables deduplication. Bellare et al. [3] formalize this primitive as message-locked encryption, and explore its application in space-efficient secure outsourced storage In the baselineapproach the convergent encryption is based on a layered approach. In this the user key is encrypted by using a master key and it is stored along with the data in cloud. Baseline approach [6] suffers critical problems. It is inefficient in the aspect that it generates enormous number of keys with increasing number of users.

Message-Locked Encryption (MLE) [3] |so named because the message is locked, as it were, under itself| with the goal of providing an encryption primitive that provably enables secure deduplication. The key generation algorithm of an MLE scheme K maps a message M to a key K. The encryption algorithm E takes input the key K and a message M and produces a ciphertext C. The decryption algorithm D allows recovery of M from C given the key K. The tagging algorithm T maps the ciphertext C to a tag T used by the server to detect duplicates. (Tag correctness requires that tags corresponding to messages M1,M2 are likely to be the same iff M1,M2 are the same.) . All algorithms may depend on a parameter P but the latter is public and common to all parties including the adversary, and thus is not a key. Any MLE scheme enables deduplication of ciphertexts. CE is captured by our syntax as the MLE scheme that lets K = H(M), C = E(K,M) and tag T = H(C). MLE is trivially achieved by letting the key K equal the message M. (Set C = T = " to the empty string and have decryption simply return the key.) This degenerate solution is however useless for deduplication since the client stores as K the entire file and no storage savings result. We rule it out by requiring that keys be shorter than messages, ideally keys are of a fixed, short length.

Fortunately there are lot of method for image compression. JPEG compression is a widely used form of lossy image compression that centers around the Discrete Cosine Transform[4].Due to the increasing requirements for transmission of images in computer, mobile environments, the research in the field of image compression has increased significantly. Image compression plays a crucial role in digital image processing, it is also very important for efficient transmission and storage of images. When we compute the number of bits per image resulting from typical sampling rates and quantization methods, we find that Image compression is needed. Therefore development of efficient techniques for image compression has become necessary[8].Three closely connected components form a typical lossy image compression system, they are (a) Source Encoder (b) Quantizer and (c) Entropy Encoder. (a) Source Encoder (or Linear Transformer) It is aimed at decorrelating the input signal by transforming its representation in which the set of data values is sparse, thereby compacting the information content of the signal into smaller number of coefficients. a variety of linear transforms have been developed such as Discrete Cosine Transform (DCT), Discrete wavelet Transform (DWT), Discrete Fourier Transform (DFT). (b) Quantizer: Aquantizer aims at reducing the number of bits needed to store transformed coefficients by reducing the precision of those values. Quantization performs on each individual coefficient i.e. Scalar Quantization (SQ) or it performs on a group of coefficients together i.e. Vector Quantization (VQ). (c) Entropy Coding Entropy encoding removes redundancy by removing repeated bit patterns in the output of the Quantizer. the most common entropy coders are the Huffman Coding, Arithmetic Coding, Run Length Encoding (RLE) and Lempel-Ziv (LZ) algorithm.

III. PROPOSED METHOD

A. *System model*

First we develop a data outsourcing model by using Dekey. It consist of three entities. User, Storage- cloud Service Provider (S-CSP), Keymanagement Cloud Service Provider (KMCSP).

User: it is an entity who wish to outsource the data to the cloud and retrieve it later. To save upload bandwidth theuser only upload unique data.
S-CSP: Provides the data outsourcing service and stores the data on behalf of the user.
KMCSP: Provide keymanagement by maintaining the convergent keys.

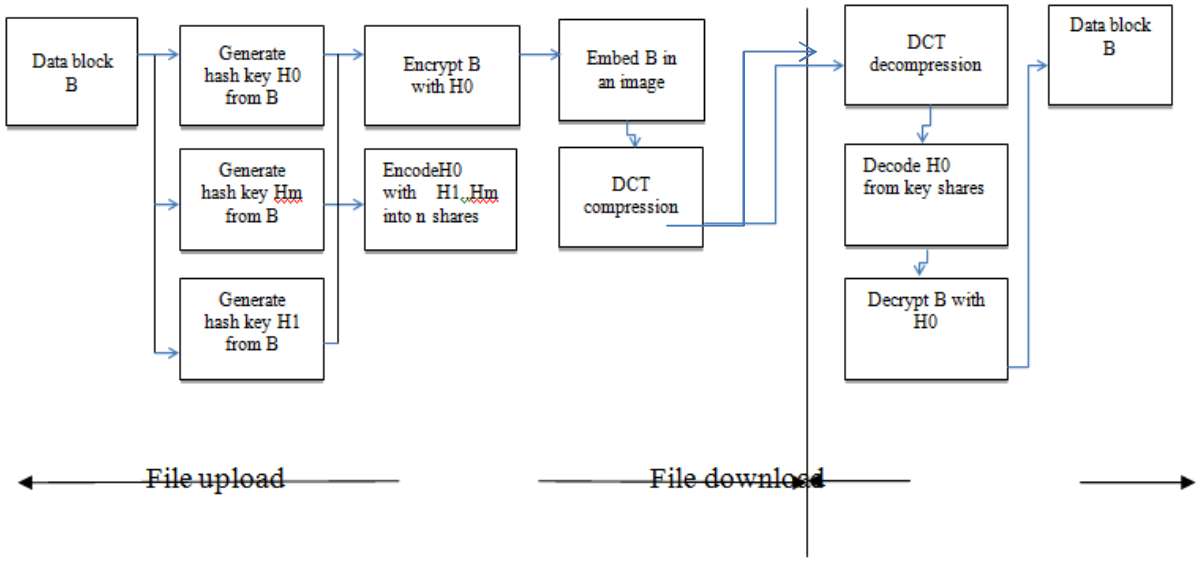
In this we specify two types of deduplication. File level deduplication that identifies duplicate files and block level deduplication in which the file is divided into fixed size blocks and avoid uploading of redundant blocks[5].

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016



IV. PSEUDO CODE

A.File Upload

To upload a file F first file level deduplication is to be performed.

Step 1: On input file F, the user computes and sends the block tags $T(B_i)=TagGen_{CE}(B_i)$ and file tag ($TagGen_{CE}$) to the S-CSP.

Step 2: Upon receiving, the S-CSP checks whether there exists the same tag on the S-CSP. If so, the S-CSP replies the user with a response ‘‘file duplicate,’’ or ‘‘no file duplicate’’otherwise.

Step 3: Upon receiving results for a block Bi returned from KM-CSPs, if it is a valid pointer, the user stores it locally; otherwise the user computes the secret shares K1,K2;then sends the share Ki and $Tj(B_i)= TagGen_{CE}(Bi,j)$ to the j-th KM-CSP for j = 1; 2; . . . ; n

Step 4: Upon receiving Kij and $Tj(B_i)$, the j-th KM-CSP stores them and sends back the pointer for Kij to the user for future access.

Step 5: Embed the data in an image. Compress the image using DCT image compression.

Step 6: Stores the data to the cloud.

B. File Download

To decrypt the data first the user has to obtain the key shares from KMCSP and recover the image from S-CSP and extract it. After extracting it decrypt the cipher text using the obtained key.

V. PERFORMANCE ANALYSIS

The performance can be analyzed on the basis of security it provides. Dekey is designed to solve the key management problem in secure deduplication where the files have been encrypted by utilizing convergent encryption and then embedding in an image.

VI. CONCLUSION AND FUTURE WORK

We propose Dekey, an efficient and reliable convergent key management scheme for secure deduplication. Dekey applies deduplication among convergent keys and distributes convergent key shares across multiple key servers, while preserving semantic security of convergent keys and confidentiality of outsourced data. We implement Dekey using visual cryptography and demonstrate that it incurs small encoding/decoding overhead compared to the network

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

transmission overhead in the regular upload/download operations. To enforce more security data is embedded in an image. To reduce storage space DCT compression is used.

REFERENCES

1. Jin Li, Xiaofeng Chen, Mingqiang Li, Jingwei Li, Patrick P.C. Lee, and Wenjing Lou Secure Deduplication with Efficient and Reliable Convergent Key Management IEEE transactions on parallel and distributed systems, vol. 25, no. 6, june 2014
2. P. Anderson and L. Zhang, ''Fast and Secure Laptop Backups with Encrypted De- Duplication,'' in Proc. USENIX LISA, 2010, pp. 1-8.
3. M. Bellare, S. Keelveedhi, and T. Ristenpart, ''Message-Locked Encryption and Secure Deduplication,'' in Proc. IACR Cryptology ePrint Archive, 2012, pp. 296-3122012:631.
4. Ken Cabeen and Peter Gent," Image Compression And Discrete Cosine Transform", Math 45 college of REDWOODS.S. Kamara and K. Lauter, ''Cryptographic Cloud Storage,'' in Proc. Financial Cryptography: Workshop Real-Protocols Standardization, 2010, pp. 136-149
5. M. Mulazzani, S. Schrittwieser, M. Leithner, M. Huber, and E. Weippl, ''Dark Clouds on the Horizon: Using Cloud Storage as Attack
6. Vector and Online Slack Space,'' in Proc. USENIX Security, 2011, p. 5.
7. M.W. Storer, K. Greenan, D.D.E. Long, and E.L. Miller, ''Secure Data Deduplication,'' in Proc. StorageSS, 2008, pp. 1-10.
8. A.M.Raid1, W.M.Khedr2, M. A. El-dosuky1 and Wesam Ahmed1"Jpeg Image Compression Using Discrete CosineTransform - A
9. Survey.International Journal of Computer Science & Engineering Survey (IJCSES) Vol.5, No.2, April 2014.
10. Mark W. Storer Kevin Greenan Darrell D. E. Long Ethan L. Miller "Secure Data Deduplication" Storage Systems Research Center
11. University of California, Santa Cruz

BIOGRAPHY

SOOFIYA M.S is a MTECH student in the Computer Science and Engineering Department, Marian Engineering College , Kerala University, Trivandrum, kerala.She received BTECHin Computer Science and Engineering degree in 2005 fromYounus college of Engineering, Kollam, Kerala, India.

SREETHA V KUMAR ASST Professor in the Computer Science and Engineering Department, Marian Engineering College , Kerala University,Trivandrum,kerala.She received MTECH from MS university in 2012.