International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 12, December 2015
Available at http://internationaljournalofresearch.org

A Composition of Different Elements on Cloud for Secure & Authorized Data Deduplication

Somavarapu Vamsi[1]& S.Vasu[2]

[1]m-Tech Dept. Of C.S.E (C.S) Sri Venkateswara College Of Engineering & Technology (Autonomous) Chittoor- 517127, Andhra Pradesh
Mail Id: -vamsidhar100@gmail.com

[2]Associate professor dept. of C.S.E (C.S) Sri Venkateswara College Of Engineering & Technology (Autonomous) Chittoor- 517127, Andhra Pradesh
Mail Id: -vasu.cs513@gmail.com

Abstract-

Here Hybrid Cloud approach is nothing but the combination of Public Cloud and Private Cloud. When user wants to store any file in Public Cloud, first he send a request to Private Cloud for File Token (hash value of file + user attributes) Then Private Cloud generates File Token and gives to User. After getting the File Token, User can send to Public Cloud for Deduplication, if File Token already is there then Public Cloud check Owner of the File, if he authenticated person then Public cloud sends a existing File reference id to user. So like that i can achieve Deduplication in Cloud. Suppose, if File Token is not available in Public Cloud then User can be perform convergent encryption and store the file in Public Cloud. As more private users outsource their data to cloud storage providers, recent data breach incidents make encryption a progressively prominent demand. Unfortunately, semantically secure encryption systems render various cost-efficient storage optimization proficiencies, such as data deduplication, ineffective.With the continuous and exponential functions growing of the number of users and the size of their data, data deduplication becomes more and more anessential for cloud storage providers (CSP). By storing a unique copy of duplicate data, cloud providers greatly mitigate their storage and data transversal costs. The advantages of deduplication unfortunately come with a high cost in terms of new security and privacy challenges. To protect the confidentiality of sensitive data while supporting deduplication, the convergent encryption proficiency has been proposed to encrypt the data before outsourcing. This traditional convergent encryption will be insecure for inevitablefile. To avoid the deterministic key generation we use of public-key encryption gives more flexibility for our application. The file F is encrypted with convergentkeyk, while k will be encrypted with Private key PK. In this way, cloud server cannot decrypt the ciphertext. We show that the overhead introduced by these new elements are minimal and does not impact the overall storage and computational costs.

Index Terms—Deduplication; confidentiality; Convergent Encryption Technique; Public-key Encryption; Cryptosystem

1. Introduction

In cloud computing, deduplication has been a well-kenned technique and has magnetized more and more attention recently. Information deduplication is a particularized information compression technique for eliminating duplicate replicas of reiterating data in storage. The technique is utilized to ameliorate storage utilization and can withal be applied to network data transfers to mitigate the number of bytes that must be sent. In lieu of keeping multiple data copies with the same content, deduplication excretes redundant information by keeping only one physical copy and referring other redundant information to that copy. Deduplication can take lay at either ye file level. For file level deduplication, it eliminates duplicate facsimiles of the same file. Deduplication can adscititiously

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 12, December 2015
Available at http://internationaljournalofresearch.org

take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files.

Albeit data deduplication brings a plethora of benefits, security and privacy concerns arise as utilizers sensitive information are sensitive to both insider and outsider attacks .Traditional encryption, while providing information confidentiality, is uncongenial with informationdeduplication. Concretely, traditional encryption requires different users to encrypt their information with their possess keys. Thus, identical data replicas of different users will extend to dissimilarciphertexts, making deduplication infeasible. Convergent encryption has been proposed to enforce data confidentiality while building deduplication feasible. It encrypts/decrypts ainformation copy with a convergent key, which is obtained by calculating the hash value of the message of the information copy. After key generation and information encryption, users hold the keys plus send the ciphertext to ye cloud. Since the encryption operation is settled and is deduced from the data content, identical l data copies will engender the same convergent key and hence the same ciphertext. To avert wildcat access, a insure proof of ownership protocol is withal needed to provide the proof that the utilizer indeed owns ye same file when a duplicate is detected. After the proof, subsequent users with the same file will be allowed foranarrowfrom the server less needing to upload the like file. A utilizer can download the encrypted file with the arrow from the server, which can alone be decrypted by the corresponding data proprietors on their convergent keys. Thus, convergent encryption sanctions the cloud to perform deduplication on the ciphertexts and the proof of ownership obviates the unauthorized utilizer to access the file.

2. Related Work

Hybrid cloud can be built utilizing any technology it changes granting to unlike vendors. Key constituents In many of the positions, effectuation of the hybrid cloud has a comptroller that will hold track of all placements of private and public clouds, IP address, servers and other resources that can run systems efficiently.

2.1Convergent encryption. Convergent encryption [1],[2] provides data confidentiality in deduplication. A data owner derives a convergent key from eachoriginal data copy and ciphers the data copy with theconvergent key. In addition, the user also derives a *tag*for the data copy, such that using ofthetag can be detectduplicates. Here, we assume that the tag correctnessproperty [1] holds, i.e., if two their tags are same,thenthedata copies are same. To detect duplicates, theuser first sends the tag to the server side to verifyiftheindistinguishable copy has been already stored. Note thatboth the convergent key and the tag are independentlydeducted, and the tag will not be used to deduce theconvergent key and compromise data privations.Both the encrypted data copy and its matchingtagcan be stored on the server side.

2.2Public Key Cryptography

Public-key cryptography [3]is a radical departure from all that has gone before. However, public-key cryptographyis based on mathematical functions and isasymmetrical in nature, affecting the use of two keys, as opposed to conventional single key encryption. Several thoughts are held about p-k:

1. That Public Key Cryptography is more secure from cryptanalysis than conventional encryption. In fact the certificate of any system depends on computational work and key length the involved in breaking the cipher.

2. That Public-key cryptography encryption has superseded single key encryption. This is unlikely due to the increased processing power required.

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 12, December 2015
Available at http://internationaljournalofresearch.org

3. That key directionwasfiddling with public key cryptography, this is not correct.

The concept of Asymmetric cryptography evolved from an attempt to solve two problems, the development of digital signatures and key distribution. In conventional encryption for encryption and decryption the same key is used. This is not a necessary condition. Instead it is possible to develop a crypto graphical system that trusts on one key for encryption and a different but associated key for decryption. It is computationally infeasible to determine the decryption key given only cognition of the algorithm and the encryption key. The P-K process follows the Steps.

1. Each system generates a pair of keys.
2. Each system publishes its encryption key and keeping its companion key private.
3. If A wishes to send a message to B it encrypts the message using public key of B's.
4. When B receives the message, it decrypts the message using its private key. Bonlydecrypt the message knows with its private key.

3. Problem Statement

Though the above solution supports the derivativeprivilege duplicate, it is inherently subject to bruteforce attacks launched by the public cloud server, whichcanretrieve files accruing into a known set. More specifically, knowing that the target file space underlyinga given ciphertext$C$ is drawn from a message space$S= \{F1, \cdots, Fn\}$ of size n, the public cloud servercan recover F laterat the most n offline encryptions. Thatis, for each $i = 1, \cdots, n$, it simply encrypts Fi to getaciphertext denoted by Ci. sothat$C= Ci$, it means thattherudimentary file is Fi. Security is thus only possiblewhen such a message is unpredictable. This traditionalconvergent encryption will be insecure for predictablefile.We design and implement a new system which coulddefend the protection for predicatable message.

The *mainidea*of our technique is that the novel encryption keypropagation algorithm. For simpleness, we will use thehash functions to define the tag generation processand convergent keys in this part. In traditional convergent encryption[4], to support duplicate check, the keyis derived from the file F by applyingsomehash function $kF= H(F)$. To avoid the deterministic keygeneration, the encryption key $kF$for file F in our schemewill be engendered with the aid of the private key cloudserver with privilege key kp. The encryption key can belooked at as the figure of $kF,p= H0(H(F), kp) \oplus H2(F)$,Where H0, H and $H2$ are all hash methods. The file F is encrypted with another key k, while $k$will be encrypted with kF,p. In this system, both the privatecloud server and CSP cannot decrypt the ciphertext.Furthermore, it is semantically secure to the CSP basedon the protection of symmetric encryption. For CSP, if thefile is unpredictable, then it is semantically secure too.

File Recovering:

Auserwants to download a file F.The user first uses his key$kF,pj$to decrypt $Ck,pj$and obtain k. Then the user uses$k$to retrieve the original file F.

4. System Model

Aiming at allowing for deduplicatedstorehouse, wesuggest the SeccureCloud system [5]. In the SeccureCloud system, wehave three entities:

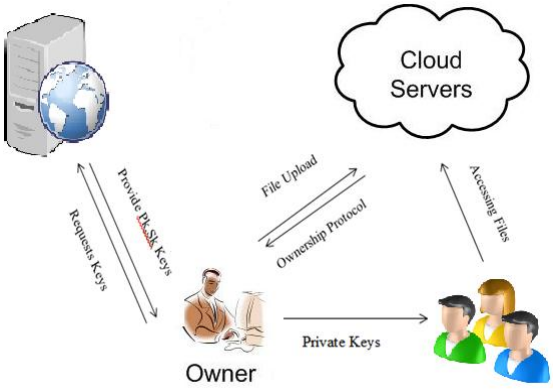


Fig-1: Secure Cloud Server

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 12, December 2015
Available at http://internationaljournalofresearch.org

DataOwner: DataOwnerperformsthe duplicate check with the cloud server to confirmif such a file is stored in cloud storage or not beforeuploading a file. If there is a duplicate, another protocolcalled Ownership will be run between the Data Owner and the cloud storage server. If it is passed, the Data Owner is authorized to access this stored file without uploadingthe file. Otherwise, file will be stored in cloud storage server. Data Owner runs the deduplication prove by sending hash value of thefileHash(F) to the cloud server.

DataUser: Data users are some peopleswhoauthorized the identity authentication and access to theData outsourced by the data owner. Notice that, theshared data can only be accessed by the authorizedusers during its authorization period.

Cloud Server: It contains almost unlimitedstorage space which is able to store and manageall the data or files in the system. Other entitieswith limited storage space can store their data tothe cloud servers.

KeyServer: It is a Key reference serverwithout any interaction with other entities involvedin the system. It is responsible for releases the Keys for Data Owner.

Ownership Protocol: It is an interactive protocol [5] initialized at the cloud server for verifying that the Data Owner exactly owns a claimed file. This protocol is typically triggeredalong with file uploading protocol to prevent the leakage ofside channel information. On the contrast to integrity protocol, in Ownership Protocol the cloud server works as verifier, while the Data Owner plays the role of prover. In this System cloud server expecting Private Keysfrom Data Owner for Ownership of the file then the client responds withthe proof for file ownership, and cloud server finallyverifies the validity of proof.

5. Implementation

We enforce a prototype of the proposed authorizeddeduplicationsystem,our implementation of the **Data Owner**provides the following function calls to support deduplication along the file upload process.

Setup (1^λ, n):performedby the data owner to arrangeanaccount on an *untrusted* server. On input a securitylevel parameter 1^λ and it outturns the public systemargumentparam, which is omitted from the inputof the other algorithms for brevity.

KeyGen:performed by the data owner to randomlygenerate a public/master-secret key pair (pk,msk).

FileUpload: It encrypts the File withConvergent Encryption applying 256-bit Blowfishalgorithmwhere the convergent key is from MD-5 Hashing of the fileand before uploading the file in to the cloud server, server will be perform duplicate check with help of hashing of the file.

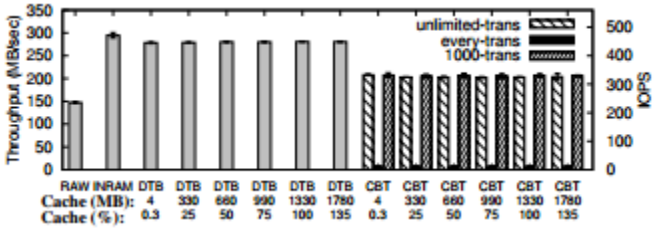
Key-Encryption: To preventing lose of Convergent Encryption Data Owner will encrypt the
Convergent key with public-key pk,and send the master-secret key to Users.

FileDownload: Data Users can access the file which is shared by Data Owner. Here User can before decrypt the file,first he should decrypt the convergent key with the master-secret key which is shared by Data Owner to User by secure channels (i.e email).

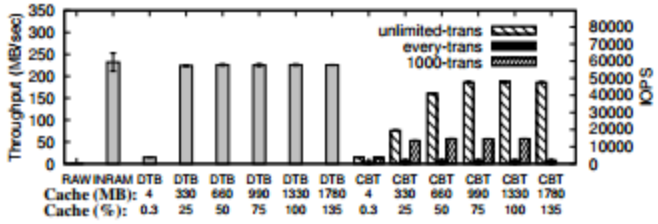
6. Experimental Results

Sequential and random write throughput for the All-duplicates dataset. Results are for the raw device and for Dmdedup with different backends and cache sizes. For the CBT backend, we varied the transaction size: unlimited, every I/O, and 1,000 writes**.**

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 12, December 2015
Available at http://internationaljournalofresearch.org



(a) Sequential write, 640KB I/O size



(b) Random write, 4KB I/O size

This dataset is on the other end of the deduplication ratio scope and all writes contain exactly the same data. Thus, after the first write, nothing needs to be written to the data device. As a result, Dmdedup outperforms the raw device by 1.4–2× for the sequential workload in all forms except CBT with per-write transactions metadata writes induced by each user write. Write amplification is lower for All-duplicates than for Unique because the hash index is not updated in the former case. The throughput does not depend on the cache size here because the hash index contains onlyone entry and fits even in the 4MB cache. The LBN mapping is accessed sequentially, so a single I/O brings in many cache entries that are immediately reaccessed. For random workloads (Figure 6(b)), Dmdedup improves performance even further: 2–140× compared to the raw device. In fact, random writes to a disk drive are so slow that deduplicating them boosts overall performance. But unlike the sequential case, the DTB backend with a 4MB cache performs poorly for random writes because LBNs are accessed randomly and 4MB is not enough to hold the entire LBN mapping. For all other cache sizes, the LBN mapping fits in RAM and performance was thus not impacted by the cache size.

7. Conclusion

In this paper, the thought of information deduplication was proposed to ensure the information security by including differential power of Data Owner in the copy check. In cloud server our information are safely store in scrambled format,and additionally in Key server our key is store with individual document.

The presentation of a couple of beginning deduplication advancements invigorating endorsed copy duplicate in crossover cloud structural engineering, in that the copy check tokens of records are induced by the private cloud server having private keys. Security check shows that the systems are secure with respect to insider and pariah attacks point by point in the proposed security model. As an issue confirmation of origination, the created model of the proposed authorized copy duplicate check strategy and tried the model. That demonstrated the endorsed copy duplicate check system experience least overhead looking at focalized encryption and information exchange.

8. References

[1]M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked
encryption and secure deduplication. In EUROCRYPT, pages 296–
312, 2013.

[2] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer.
Reclaiming space from duplicate files in a serverlessdistributedfile system. In ICDCS, pages 617–624, 2002

[3]http://www.facweb.iitkgp.ernet.in/~sourav/PublicKeyCrypto.pdf

[4],http://www.ijcea.com/wpcontent/uploads/2014/11/03Gaurav-Kakariya-et-al..pdf

[5] Bellare, Mihir, ChanathipNamprempre, and Gregory Neven. "Security proofs for identity-based identification and signature schemes." Journal of Cryptology 22.1 (2009): 1-61.

[6] M. Bellare, S. Keelveedhi, and T. Ristenpart.Dupless: Serveraided encryption for deduplicated storage. In *USENIX Security Symposium*, 2013.

[7] K. Zhang, X. Zhou, Y. Chen, X.Wang, and Y. Ruan.Sedic: privacyaware data intensive computing on hybrid clouds. In *Proceedings of the 18th ACM conference on Computer and communications security*, CCS'11, pages 515–526, New York, NY, USA, 2011. ACM.

[8] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg.Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, *ACM Conference on Computer and Communications Security*, pages 491–500.
ACM, 2011.

[9] Bugiel, Sven, et al. "Twin clouds: Secure cloud computing with low latency." Communications and Multimedia Security.Springer Berlin Heidelberg, 2011.

[10] Anderson, Paul, and Le Zhang. "Fast and Secure Laptop Backups with Encrypted De-duplication." LISA. 2010.

[11] Bellare, Mihir, SriramKeelveedhi, and Thomas Ristenpart. "DupLESS: server-aided encryption for deduplicated storage." Proceedings of the 22nd USENIX conference on Security.USENIX Association, 2013.

[12] Bellare, Mihir, SriramKeelveedhi, and Thomas Ristenpart. "Message-locked encryption and secure deduplication."Advances in Cryptology–EUROCRYPT 2013.Springer Berlin Heidelberg, 2013.296-312.

[13]. J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl.A secure data deduplication scheme for cloud storage.In Technical Report, 2013

[14] J. Yuan and S. Yu. Secure and constant cost public cloudStorage auditing with deduplication. IACR CryptologyePrint Archive, 2013:149, 2013