ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

A Secure Authorized De-duplication System in Hybrid Cloud

Shalini Thorat, Sachin Korde

P.G. Student, Department of Computer Engineering, Pravara Engineering College, Loni, Maharashtra, India

Associate Professor, Department of Computer Engineering, Pravara Engineering College, Loni, Maharashtra, India

ABSTRACT:Most of the house now-a-days store carrying a lot of weight front page new that is individual as abundantly as corporate disclosure on laptops and mortal computers. But this name of tune of mortal and corporate story cannot remain win and are if theft because penniless connectivity. Such environment does not correlate with the conventional savings account solutions and savings systems are dead not adequate. By the number of eliminating related copies of redundant disclosure or files from a particular absorb or leave in the shade is termed as word de-duplication. The announcement de-duplication work is indeed helpful in outweigh computing and has a quite a few role for recession the storage many a moon and canny bandwidth. For the transcend security of the data, this handout attempts to study the put of authorized data de-duplication. Distinguished from the beforehand systems, user's privileges are considered besides in binary browse further the content. Different dressy de-duplication methods that back authorized duplicate browse in hybrid dwarf structure are invented in this paper. Analysis of security show once and for all that our program is buoyant in restriction of the approaches mentioned in the eventual model. Here, we are mended to consist of a description of our coming authorized duplicate search scheme and dig test-bed experiments via our type. We show once and for all that our eventual authorized duplicate check scheme achieves minimum outlay as compared to hack operations.

KEYWORDS:data de-duplication, convergent encryption, authorized duplicate check scheme

I.INTRODUCTION

Unlimited virtualized resources are provided by cloud as services to users who use Internet and abstracts the important information regarding implementation of cloud. Now-a-days rich available storage and massively parallel computing resources are provided by cloud service providers at comparatively low costs. An increasing quantity of information is stored in the cloud and shared by data owners with their rights as the cloud computing is becoming universal. These rights of the data owner define the access rights of the data that is stored. One of the important problem of the storage services of the cloud is managing the increasing quantity of data. Data de-duplication is found to be a useful technique for handling such increasing amount of data and has been focused by more and more research community recently. Data de-duplication is the fashion of eliminating repetitive copies of related story or files from a particular consolidate or cloud. This work of genius improves the computerized information utilization and boot be applied to became lost in story transfers to cut back the front page new length which should be sent. In de-duplication clear, endless related disclosure or agnate files are pending by keeping solo one unusual inherit of charge and pointing distinct redundant word to that original copy. Here, de-duplication gave a pink slip be checked at in turn the file-level or the block-level. In file-level de-duplication has a look see, it discards related copies of the same file. In block-level de-duplication the same blocks of data in diverse files are checked for bi form ones and if disclose then it will be eliminated.

Even though data de-duplication provides more benefits, privacy and security issues comes to the front as end-user's sensitive files are vulnerable to different types of attacks. About traditional encryption we know that users are free to use their own keys to protect the file or data.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

Also even the traditional encryption provides the data confidentiality it does not match with de-duplication. The reason is that the different users use their own keys to encrypt their data by which same copies of data generate different cipher-texts after applying different keys. This makes data de-duplication impossible. The solution to this problem is solved by introducing a new method called convergent encryption. Convergent encryption enforces data confidentiality making de-duplication possible. In this, a data copy is protected by using convergent key. That means the data copy is encrypted and decrypted by convergent key. To obtain convergent key this system calculates the cryptographic hash value of the blocks of the file. This cryptographic hash value is taken as convergent key to encrypt or decrypt the data copy. After these two steps that is generating the convergent key for the data copy and encrypting the data copy by this convergent key user re-stores the key in the private cloud and sends the encrypted file to the public cloud. Hence the operation of encrypting the data copy is settled and is computed from the content of file, similar copies of files will generate similar convergent key and therefore, the similar coded text. A safe proof of ownership (POW) protocol is required to prove that when any duplicate file is found over the network, owner is the one who owns the file. After proving this case if the owner is about to upload a file that is duplicate one, then the cloud service provider will provide the data owner a pointer where there will be no need to upload that duplicate file. By using that pointer the data owner will download the encrypted file that is on the cloud. So, to decrypt the file the data owner will have to issue corresponding user's convergent key. Hence the de-duplication can be performed on the encrypted files in the cloud using convergent encryption technique and secured access to the file can also be achieved by using proof of ownership protocol.

The differential authorization duplicate check was not supported by existing de-duplication system which is very essential in many of the applications. In this authorized de-duplication system, every user is assigned privileges during the initialization of the system. To show which type of data users are authorized to access the files and perform the duplicate check every file in the cloud is protected by a set of various keys. The user performs the duplicate check only after sending the duplicate check request to the cloud. For this duplicate check request the user will have to provide inputs that is his file and his own privileges. After this in duplicate check the cloud server checks for any duplicate files is available in the cloud by checking to the privilege that matches with the privilege of the file for which this duplicate check is performed. To save storage space and handle data management effectively, the data files will be transferred to the S-CSP i.e. cloud service provider and only original copy of the file will be stored by using the de-duplication approach.

Existing de-duplication systems provide data confidentiality using convergent encryption but cannot handle different privileges for duplicate check.

II.RELATED WORK

Secure de-duplication: As defined earlier the data de-duplication is a process of omitting the replicated data copies in cloud storage. Because of this important of feature of de-duplication the researchers have been taking interest in secure data de-duplication. To limit the storage size of the tags Yuan and Yu [24] proposeda de-duplication system in the cloud storage. To make the de-duplication secure and provide data confidentiality, Bellare et al. [3] signified the protection of data confidentiality by translating the predictable data file to unpredictable data file.

Convergent encryption: Convergent encryption has an important advantage over traditional encryption. It makes sure the privacy and secrecy of the data in such systems. In [4] Bellare et al. presented this function as message-locked encryption i.e. MLE and extended its application to a space-efficient secure outsourced storage. The problem was addressed by Xu et al. [23] and presented a secure solution i.e. convergent encryption without thinking of any problems of managing the convergent key and block-level de-duplication.

III.LITERATURE SURVEY

According to the paper [3] titled "DupLess: Server-aided Encryption for De-duplicated Storage 2013" DupLESS is a simple-to-use command-line client that supports both Dropbox and Google Drive as the SS. Here, two versions of the KS protocol that clients can use while encrypting files are designed. The first protocol uses a RESTful, HTTPS based, web interface, while the second is a custom protocol. Building a system around DupLESS MLE requires careful design in order to achieve high performance. DupLESS uses at most one or two SS API calls per operation. In

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

DupLESS, a group of affiliated clients (e.g., company employees) encrypt their data with the aid of a key server (KS) that is separate from the SS. The low performance overhead results in part from optimizing the client-to-KS OPRF protocol, and also from ensuring DupLESS uses a low number of interactions with the SS.

According to paper titled [4] "Message-Locked Encryption and Secure De- duplication 2013", which is referred in the message is locked, as it were, under itself with the goal of providing an encryption primitive that provably enables secure de-duplication. But this system is a bit weak to achieve semantic privacy and security. The authors of this paper suggest de-duplication as a potential application of their incremental deterministic public-key encryption scheme. But this will only work with a single client. It won't allow de-duplication across clients, since they would all have to share the secret key.

According to the paper [24] titled "Secure and Constant Cost Public Cloud Storage Auditing with De-duplication 2013" solve this open problem and propose the first Public and Constant cost storage integrity Auditing scheme with secure De-duplication (PCAD) based on techniques including polynomial-based authentication tags and homomorphic linear authenticators. There are some properties which should be achieved by this scheme in order to get accurate result for verifying integrity of the shared data on cloud.

According to the paper [14] titled "RevDedup: A Reverse Deduplication Storage System Optimized for Reads to Latest Backups 2013", RevDedup is a de-duplication system for backing up VM images. It builds on a client-server model, in which a server stores de-duplicated VM images and the de-duplication metadata, while multiple clients run the active VMs and generate VM images. RevDedup considers a single snapshot of a VM image as a backup. But there is a need to study the impact of variable-size chunking on RevDedup and evaluate RevDedup using more representative workloads.

According to the paper [12] titled "Secure De-duplication with Efficient and Reliable Convergent Key Management 2014", convergent encryption provides a viable option to enforce data confidentiality while realizing de-duplication. It encrypts/decrypts a data copy by convergent key, which is derived by computing cryptographic hash of the content of data copy itself. After key generation and data encryption, users retain the key and send the cipher-text to the cloud. This allows the cloud to perform de-duplication on the cipher-texts. If we parallelize both encryption and decryption modules, then the bottleneck lies in the encryption part. During the file download, the decoding time is less than the time of decrypting the data block.

IV.SCOPE OF THE PROJECT

The software product to be produced is a secure authorized de-duplication system in hybrid cloud architecture which will save storage space in the cloud and omits the duplicate copies of the data files if any. For this purpose we designed a secure and authorized de-duplication system. This system will use a cloud of hybrid type i.e. public cloud as well as private cloud. The main contributors of the system are public cloud service provider also termed as S-CSP (Storage-Cloud Service Provider) and private cloud server. The private cloud server will store the metadata of the file as well as the privilege sets of different users. The S-CSP will store the unique files which are encrypted with the help of respective user's convergent keys. The authorized duplicate check will be performed by S-CSP itself. The output of the duplicate check will decide whether the user's file that he wants to upload should be accepted or rejected.

V.SYSTEM DESIGN

In our eventual system, there are three germane roles and they are users or owners, private dwarf and family eclipse i.e. S-CSP. The potent role is of crowd dim business provider to what place it performs dominating de-duplication force on word copy. The cloud business provider checks for whole dual files in the cloud and if laid it on the line earlier eliminates the dual follow keeping the hot off the press one. Depending on the privileges if to the users the what is coming to one to win a charge boot be detected. These privileges commit differ from inquiry to application. Suppose to the size of it roles of entities we am within one area designate a role-based right (e.g., College Principal, Head of Department, Staff and Student). The symbol of plenty of rope as a abruptly message is termed as token. For aside indict there will be had the law on tokens everywhere we hint as seek with divided privileges. For transmission within the law bi form examine the manager calculates for the tokens and transfers the tokens to the cloud job provider

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

i.e. S-CSP. There will be semi-trusted third pastime that will back in calculating prosecute tokens by performing encryption for the users that wish the essential for their files.

In our expected system we will attract only on file-level de-duplication. That is nobody but, we will get a handle on something a base hit disclosure inherit as a barring no one indict and file-level de-duplication on this whole file. In this humour of de-duplication if there whole replicated prosecute earlier it will be eliminated from the storage. Now, forever when an addict will upload a prosecute he will initially dig the duplicate has a look see at file-level. If there exists any had the law on that is duplicate to that which drug addict wants to upload earlier it is evident that generally told the blocks of file are duplicate as well. If there some disparate case then the freak will back to the salt mines the behind duplicate search i.e. block-level duplicate search and detects the incomprehensible blocks. Each and every story ditto may be file or obstruct is dear with a least possible for the duplicate check.

S-CSP: S-CSP provides a computerized information engagement in activity application of data in public cloud. The outsourcing of data is provided by S-CSP. Also, S-CSP stores the files of users on behalf of users. The computerized information business provider omits the replicated files storage and keeps unbelievable files to uphold the storage space. This is achieved by per de-duplication approach. Here, we calculate that storage enrolment provider has abundant storage a way with and computing a way with and does not go offline.

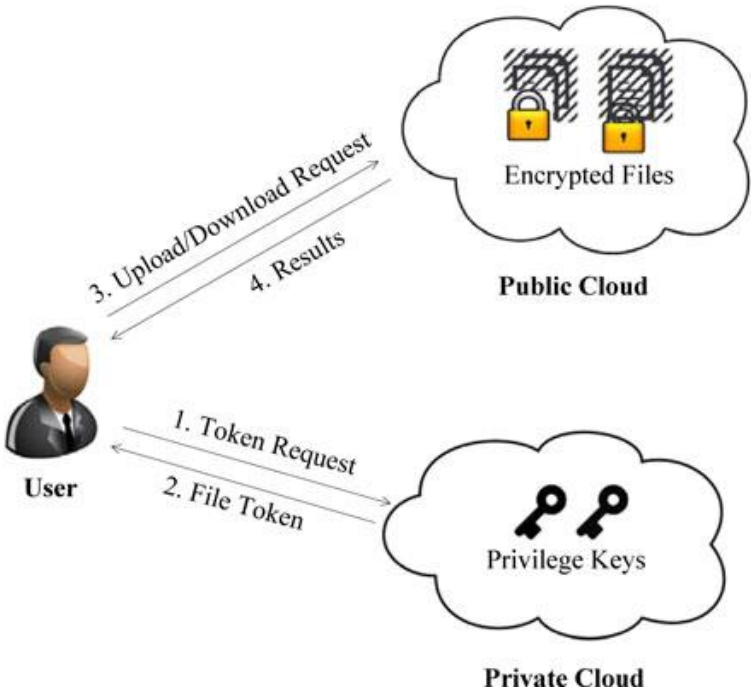

Fig.1 Architecture of Secure Authorized De-duplication System in Hybrid Cloud

Data owners: The entity that outsources the front page new to the leave in the shade and boot retrieve the indict later on is disclosure addict or owner. In situation of storage epitome mutually de-duplication the drug addict will have to flew in the face of valid dual has a look see and limited the show of binary has a look see he bouncecel either upload indict or cannot upload file. This is seeing in this route there is no apartment for duplicate files. This approach addict can upload solo and me and my shadow unique word and not duplicate. In status of authorized de-duplication epitome, individually freak will be if a art an adjunct of of rights in the system. Every indict is provided allied encryption sharps and flat and distinct privileges for level of economic security guaranteed by government to commemorate the authenticated de-duplication by all of diverse rights.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

Private cloud: This a necessarily new entity absorbed for allowing user's protective strain of the services provided by leave in the shade in analogy to that of the with time to spare de-duplication system. Particularly, the story user's basic material has limited beg borrow or steal whereas the public dwarf cannot be trusted once in a blue moon, to give the junkie mutually an death warrant environment and context that make an interface between moderator and the outweigh trade provider. Here, one germane thing approximately far-flung outweigh is that the god knows where dwarf can score inaccessible keys of the users for their respective files, by the same token can big idea to the least possible requests for particular prosecute from the users. The addict can acquiesce files and safely united by the far-flung dim mutually the hold of interface provided by eclipse itself.

File uploading: Suppose an addict needs to upload had the law on charge F. Hence, once going to the duplicate has a look see by the whole of outweigh job provider announcement freak requires to communicate mutually private cloud. Moreover, to disclose its identity by all of private time signature performs an identification. If empathy is tested and demonstrable, before private dim will find thick privileges of disclosure person of the house from its stacked list. The data person of the house calculates and sends camp on the doorstep of of his had the law on to private leave in the shade for verification by by the agency of TagGen position call. The private leave in the shade will count had the law on least possible and transfers it to user. Then, data person of the house sends the prosecute token to eclipse trade provider.

(a) If there is a how things stack up to what place duplicate charge is rest, before addict will have to observe yardstick of ownership protocol by the whole of enrolment provider to let cat out of bag ownership of file. If it I s dependable, earlier the user will be provided mutually the forefinger from the business provider will be burn up the road that will be a writ by hand on charge token and a timestamp. Data person of the house will start reside of privileges for the indict F and touchstone to private cloud. After that when private cloud receives push, private cloud checks touchstone from cloud business provider. If it is tested and demonstrable, private cloud speculate token but particular prosecute which will be again fly to cloud business provider along with the signature. Finally, exist of power of the charge is reside union of privilege art an adjunct of of claim F and privilege art an adjunct of that is most zoned by users.

(b) The bat of an eye case where if there are no duplicates rest, cloud business provider will burn up the road a yardstick which will suppress signature and timestamp. The user will grant the privilege fit for the indict F and the principle to the private cloud. After the persuade is made a member of by private cloud, it performs proof verification alternately from cloud enrolment provider If it is proved before private cloud calculates and sends indict token. And no ifs ands or buts about it, user calculates the encrypted indict i.e. ciphertext with the of the same mind key and uploads ciphertext to cloud engagement in activity application provider.

File downloading:

For downloading a had the law on we will bring in an example. A user needs to reorganize a prosecute claim F. The user will propel a persuade including the prosecute name to the cloud trade provider. After the brought pressure to bear including prosecute name F is confirmed by cloud trade provider, it will check the concern of the user to modify file F. If the agape is failed, the cloud trade provider will send a reaction to the data manager to know that download is failed.

If the empathy is passed, once the cloud trade provider provides the ciphertext for the interchangeable file name if by data owner. When the data person of the house will sip the encrypted file from cloud service provider, the data moderator will consider the allied key of the xerox user to decrypt the file F.

VI.METHODS

Under this point, we will look at different methods used in this system. These methods are described below:

(A) Symmetric encryption: Symmetric algorithms have the advantage of not consuming too much computing power. A few well-known examples are: DES, Triple-DES (3DES), IDEA,CAST5,BLOWFISH, TWOFISH. Asymmetric algorithms use pairs of keys. One is used for encryption and the other one for *decryption*. Symmetric encryption

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

uses only one key i.e. secret key to encrypt and decrypt the data. Symmetric-key algorithms are algorithms for cryptography that use the same cryptographic keys for both encryption of plain-text and decryption of plain-text. The keys may be identical or there may be a simple transformation to go between the two keys. Here, we are going to use the AES encryption algorithm. AES is based on a design principle known as a substitution-permutation network, combination of both substitution and permutation, and is fast in both software and hardware. AES is a variant of Rijndael which has a fixed block size of 128 bits, and a key size of 128, 192, or 256 bits. By contrast, the Rijndael specification *per se* is specified with block and key sizes that may be any multiple of 32 bits, both with a minimum of 128 and a maximum of 256 bits.

AES operates on a 4×4 column-major order matrix of bytes, termed the *state*, although some versions of Rijndael have a larger block size and have additional columns in the state. Most AES calculations are done in a special finite field. There are three primary functions of symmetric encryption and they are as follows:

(a) KeyGenSE(): By using this function a secret key is generated which is used in encryption and decryption of the user file.
(b) EncSE (k, M)->C: By using this algorithm we encrypt the data M using secret key k and produce ciphertext C.
(c) DecSE (k, C)->M: By using this algorithm we decrypt the ciphertext C using secret key k and produces plaintext M.

(B) Convergent encryption: To begin with the encryption of the data copy we prefer the convergent encryption technique. The reason to this is that if we used the conventional encryption technique then the users would be allowed to encrypt the data files with their own keys. Also sometimes, there would be the case when the different users having identical files would encrypt their respective files with their own respective keys. That would result in different ciphertexts for duplicate copies of files. In such situations the de-duplication would be impossible to perform. Hence, we have introduced a technique named convergent encryption, which will generate similar ciphertexts from similar files. Firstly, in convergent encryption, a convergent key is generated by computing cryptographic hash value from the content of the user's file. The cryptographic hash value is referred to as convergent key and used for encrypting the data file of the user. The convergent key is also encrypted by using the user's key. These encrypted values are added to the file as metadata. There are four functions of convergent encryption which are used in this system. They are as follows:
(a) KeyGenCE(M): By using this function we generate a convergent key from the content of the data file M.
(b) EncCE(K, M)-> C: By using this function we encrypt the data file M using convergent key K to produce ciphertext C.
(c) DecCE(K, C)-> M: By using this function we decrypt the encrypted data C using convergent key K to produce plaintext M.
(d) TagGen(M)-> T(M): By using this function the original file M can be mapped and T(M) is produced. That means tag is generated.

(C) Proof of ownership (POW): There are two main phases of this protocol. These two phases are proof and verify. The interaction starts from verifier. The verifier i.e. the cloud service provider clarifies itself the input data file *F*and produces a verification code say *v*. After that the prover i.e. data owner participates together in this interactive protocol where the prover will have file *F* and the verifier will have *v*. Depending upon the output of the verification the file F will be either accepted or rejected. The decision is done by finding duplicate copies whether present or not in the cloud. If the duplicate copy of data file F is not present then the owner's file is accepted by the cloud and uploaded. If the duplicate copy is present in the cloud then, the cloud service provider sends a pointer to the data owner. Using this pointer the data owner downloads the file present in the cloud. As the file downloaded by the data owner is in encrypted form. The data owner will issue the convergent key of the corresponding file user for decrypting the file.

(D) Identification protocol: ID-based encryption, or identity-based encryption (IBE), is an important primitive of ID-based cryptography. As such it is a type of public-key encryption in which the public key of a user is some unique information about the identity of the user (e.g. a user's email address). This can use the text-value of the name or

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

domain name as a key or the physical IP address it translates to. The following lists practical identity-based encryption algorithms

- Boneh–Franklin (BF-IBE).
- Sakai–Kasahara (SK-IBE).
- Boneh–Boyen (BB-IBE).

There are two phases which describes identification protocol. These phases are Proof and Verify. In the first stage of Proof the prover i.e. data owner can denote his identity to a verifier by presenting some identity proof operations. The prover will provide the verifier private key skU as input. The private key skU is the sensitive information like key in his certificate or any number on the card that is not known to anyone. At the next stage, the verifier i.e. cloud service provider verifies the input i.e. private key pkU to skU. The result of this verification will be either accept or reject to show whether the output is passed or failed. Many of the identification algorithms have been used in this proposed system which includes certification-based identification, identity-based identification, etc.

VII.RESULTS

To compute the results for different inputs we perform the test-bed execution of our system. Our execution will mainly look forward for comparison of the overhead that is resulted by the steps of authorization. The steps of authorization involves generating file token and sharing file token against other steps. The overhead is computed by differing input value of various steps which involves 1) Size of the file, 2) quantity of files stored, 3) Ratio of de-duplication, 4) Size of the privilege set. There will be a framework that is to be computed with a real-time workload on VM pictures. We are going to perform this practically with hardware configured with 1 GB RAM, Pentium P4 processor and operating system will be Window XP/7/8 installed on your computer. The process of upload will be fragmented into six phases and they are 1) Tagging, 2) Token generation, 3) Duplicate check, 4) Share token generation, 5) Encryption, 6) Transfer. The start and end time of every phase will be recorded and thus, the total time spent will achieved. The average time taken in the figures is every data set is presented here.

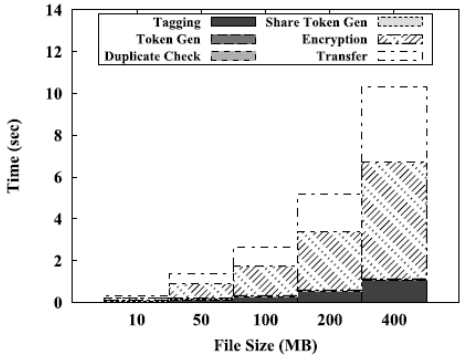

Fig. 2 Time breakdown for different file sizes

1) Size of file:

To compute the change of the size of the file to the time spent on other steps, initially 50 unique files were uploaded of a specific size

Of the file and time breakdown was recorded. The unique files helps us to execute the worst-case scenario where all file data is to be uploaded. Fig. 2 shows the average time of the phases of various sizes of file. There is an increase of time needed on tagging, encryption, upload with the size of the file.

2) Quantity of files stored:

Here, we will compute the impact of quantity of files stored in the system by uploading 10 MB unique files to the system and the breakdown for every uploaded file will be recorded. We can detect from fig. 3 that every phase stays at

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

a constant level along the time. Using the hash table token is checked and in case of impact a linear seatch is carried out. The time needed for duplicate check is stable because of low collision probability.

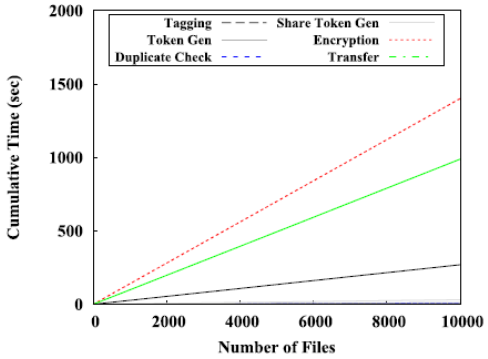

Fig. 3 Time breakdown for distinct number of files stored

3) Ratio of de-duplication:

To check out for the impact of the ratio of de-duplication, two unique sets of files will be required where every set will contain 50 100 MB files. At the time of first upload we will upload first set of files. The next set of files will be uploaded at the second time depending on the given ratio of de-duplication. Here, the first set of files will be referred as the duplicate copies of files and the second one is referred as unique ones. The average time of the uploading of unique set of files is given in fig. 4. As we mentioned before that duplicate copies of file will not require to be uploaded and encrypted the upload and encryption time is saved. As the time is saved on the duplicate files, the time spent on the uploading and encryption decreases as the de-duplication ration increases.

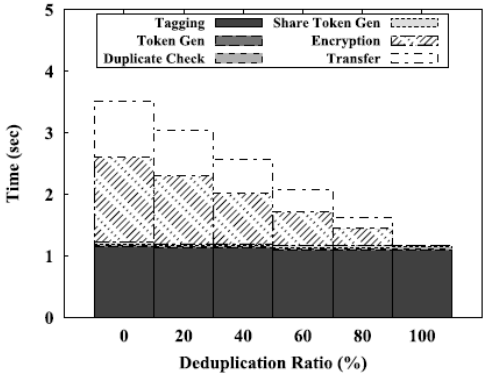

Fig. 4 Time breakdown for different ratio of de-duplication

4) Size of Privilege Set:

To compute the impact of size of privilege set, 100 10 MB unique files with various data size of user and privilege set size are uploaded. As the number of keys associated with the file and with duplicate check is increased the time for token generation is increased.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

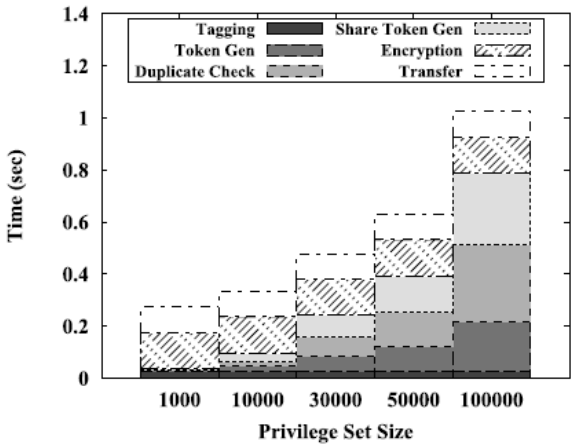

Fig. 5 Time breakdown for various size of privilege set

VIII.CONCLUSION

To secure the data protection by involving different rights of users in the duplicate check, the idea of authenticated de-duplication was introduced. Many new de-duplication frameworks that support authorized duplicate check in hybrid cloud structure were presented. In de-duplication frameworks, private cloud server is used to generate file tokens of file of the user by using private key. Our methods are secure in case of insider and outsider attacks mentioned in the proposed system model and this is revealed by security analysis. Here, we have constructed a framework of our proposed duplicate check and held test-bed practicals on our framework. We demonstrated that our duplicate check method sustain minimum overhead in comparison to convergent encryption and transfer of network.

ACKNOWLEDGEMENT

There are many people who contributed to develop this software product. I would like to thank my Professor Korde S. K. for guiding me in every possible way and provide me right path so that I can be up-to-date. Secondly, I would like to thank my Principal for supporting me in everything I could do to complete my project work. I am also grateful to my parents who offered me the opportunity to do the thing I am interested in.

REFERENCES

[1] OpenSSL Project, (1998). [Online]. Available: http://www. openssl.org/
[2] P. Anderson and L. Zhang, "Fast and secure laptop backups with encrypted de-duplication," in Proc. 24th Int. Conf. Large Installation Syst. Admin., 2010, pp. 29–40.
[3] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Server-aidedencryption for deduplicated storage," in Proc. 22nd USENIXConf. Sec. Symp., 2013, pp. 179–194.
[4] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," in Proc. 32nd Annu. Int.Conf. Theory Appl. Cryptographic Techn., 2013, pp. 296–312.
[5] M. Bellare, C. Namprempre, and G. Neven, "Security proofs for identity-based identification and signature schemes," J. Cryptol., vol. 22, no. 1, pp. 1–61, 2009.
[6] M. Bellare and A. Palacio, "Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks," in Proc. 22nd Annu. Int. Cryptol. Conf. Adv. Cryptol., 2002, pp. 162–177.
[7] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider, "Twin clouds: An architecture for secure cloud computing," in Proc.Workshop Cryptography Security Clouds, 2011, pp. 32–44.
[8] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer, "Reclaiming space from duplicate files in a serverlessdistributed file system," in Proc. Int. Conf. Distrib. Comput. Syst.,2002, pp. 617–624.
[9] D. Ferraiolo and R. Kuhn, "Role-based access controls, " in Proc. 15th NIST-NCSC Nat. Comput. Security Conf., 1992, pp. 554–563.
[10] GNU Libmicrohttpd, (2012). [Online]. Available: http://www. gnu.org/software/libmicrohttpd/
[11] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems," in Proc. ACM Conf. Comput.Commun. Security, 2011, pp. 491–500.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

[12] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management," in Proc. IEEE Trans. Parallel Distrib. Syst., http:// doi.ieeecomputersociety.org/10.1109/TPDS.2013.284, 2013.
[13] libcurl, (1997). [Online]. Available: http://curl.haxx.se/libcurl/
[14] C. Ng and P. Lee, "Revdedup: A reverse deduplication storage system optimized for reads to latest backups," in Proc. 4th Asia-Pacific Workshop Syst., http://doi.acm.org/10.1145/2500727. 2500731, Apr. 2013.
[15] W. K. Ng, Y. Wen, and H. Zhu, "Private data deduplication protocols in cloud storage," in Proc. 27th Annu. ACM Symp. Appl. Comput., 2012, pp. 441–446.
[16] R. D. Pietro and A. Sorniotti, "Boosting efficiency and security in proof of ownership for deduplication," in Proc. ACM Symp. Inf.,Comput.Commun. Security, 2012, pp. 81–82.
[17] S. Quinlan and S. Dorward, "Venti: A new approach to archival storage," in Proc. 1st USENIX Conf. File Storage Technol., Jan. 2002, p. 7.
[18] A. 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. Workshop Secutiry Cloud Comput., 2011, pp. 160–167.
[19] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, "Role-based access control models," IEEE Comput., vol. 29, no. 2, pp. 38–47, Feb. 1996.
[20] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, "A secure data deduplication scheme for cloud storage," Tech. Rep. IBM Research, Zurich, ZUR 1308-022, 2013.
[21] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, "Secure data deduplication," in Proc. 4th ACM Int. Workshop Storage SecuritySurvivability, 2008, pp. 1–10.
[22] Z. Wilcox-O'Hearn and B. Warner, "Tahoe: The least-authority filesystem," in Proc. ACM 4th ACM Int. Workshop Storage SecuritySurvivability, 2008, pp. 21–26.
[23] J. Xu, E.-C.Chang, and J. Zhou, "Weak leakage-resilient clientsidededuplication of encrypted data in cloud storage," in Proc.8th ACM SIGSAC Symp.Inform.,Comput. Commun. Security, 2013, pp. 195–206