ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Secure Authorized Duplication with a Hybrid Cloud Approach

Chavala Baji Babu
PG Scholar,
Department of CSE,
Sri Chundi Ranganayakulu Engineering College,
Chilakaluripet, Guntur, AP, India.

G.Mallikharjuna Rao
Assistant Professor,
Department of CSE,
Sri Chundi Ranganayakulu Engineering College,
Chilakaluripet, Guntur, AP, India.

ABSTRACT:

Data deduplication is one of important data compression techniques for eliminating duplicate copies of repeating data, and has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To protect the confidentiality of sensitive data while supporting deduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this paper makes the first attempt to formally address the problem of authorized datam deduplication. Different from traditional deduplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. We also present several new deduplication constructions supporting authorized duplicate check in a hybrid cloud architecture. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement a prototype of our proposed authorized duplicate check scheme and conduct testbed experiments using our prototype. We show that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations.

INTRODUCTION:

Cloud computing provides seemingly unlimited “virtual ized”resources to users as services across the wholeInternet, while hiding platform and implementation details. Today’s cloud service providers offer both highlyavailable storage and massively parallel computing resourcesat relatively low costs. As cloud computingbe comes prevalent, an increasing amount of data is beingstored in the cloud and shared by users with specifiedprivileges, which define the access rights of the storeddata. One critical challenge of cloud storage services isthe management of the ever-increasing volume of data.To make data management scalable in cloud computing,deduplication [17] has been a well-known techniqueand has attracted more and more attention recently.

Data deduplication is a specialized data compressiontechnique for eliminating duplicate copies of repeatingdata in storage. The technique is used to improve storage utilization and can also be applied to network datatransfers to reduce the number of bytes that must besent. Instead of keeping multiple data copies with thesame content, deduplication eliminates redundant databy keeping only one physical copy and referring otherredundant data to that copy. Deduplication can takeplace at either the file level or the block level. For filelevededuplication, it eliminates duplicate copies of thesame file. Deduplication can also take place at the blocklevel, which eliminates duplicate blocks of data thatoccur in non-identical files.Although data deduplication brings a lot of benefits,security and privacy concerns arise as users’ sensitivedata are susceptible to both insider and outsider attacks.Traditional encryption, while providing data confidentiality,is incompatible with data deduplication. Specifically,traditional encryption requires different users toencrypt their data with their own keys.

Thus, identicaldata copies of different users will lead to different ciphertexts,making deduplication impossible. Convergentencryption [8] has been proposed to enforce data confidentialitywhile making deduplication feasible. It encrypts/decrypts a data copy with a convergent key, whichis obtained by computing the cryptographic hash valueof the content of the data copy. After key generationand data encryption, users retain the keys and send theciphertext to the cloud. Since the encryption operation isdeterministic and is derived from the data content, identicaldata copies will generate the same convergent keyand hence the same ciphertext. To prevent unauthorizedaccess, a secure proof of ownership protocol [11] is alsoneeded to provide the proof that the user indeed ownsthe same file when a duplicate is found. After the proof,subsequent users with the same file will be provided apointer from the server without needing to upload thesame file. A user can download the encrypted file withthe pointer from the server, which can only be decryptedby the corresponding data owners with their convergentkeys.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Thus, convergent encryption allows the cloud toperform deduplication on the ciphertexts and the proofof ownership prevents the unauthorized user to accessthe file. However, previous deduplication systems cannot supportdifferential authorization duplicate check, which is importantin many applications. In such an authorizeddeduplication system, each user is issued a set of privilegesduring system initialization (in Section 3, weelaborate the definition of a privilege with examples).Each file uploaded to the cloud is also bounded by a setof privileges to specify which kind of users is allowed toperform the duplicate check and access the files. Beforesubmitting his duplicate check request for some file, theuser needs to take this file and his own privileges asinputs. The user is able to find a duplicate for this fileif and only if there is a copy of this file and a matchedprivilege stored in cloud. For example, in a company,many different privileges will be assigned to employees.

In order to save cost and efficiently management, thedata will be moved to the storage server provider (SCSP)in the public cloud with specified privileges andthe deduplication technique will be applied to store onlyone copy of the same file. Because of privacy consideration,some files will be encrypted and allowed theduplicate check by employees with specified privilegesto realize the access control. Traditional deduplicationsystems based on convergent encryption, although providingconfidentiality to some extent, do not supportthe duplicate check with differential privileges. In otherwords, no differential privileges have been consideredin the deduplication based on convergent encryptiontechnique. It seems to be contradicted if we want torealize both deduplication and differential authorizationduplicate check at the same time.

CONTRIBUTIONS:

In this paper, aiming at efficiently solving the problem ofdeduplication with differential privileges in cloud computing,we consider a hybrid cloud architecture consistingof a public cloud and a private cloud. Unlike existingdata deduplication systems, the private cloud is involvedas a proxy to allow data owner/users to securely performduplicate check with differential privileges. Such anarchitecture is practical and has attracted much attentionfrom researchers. The data owners only outsource theirdata storage by utilizing public cloud while the dataoperation is managed in private cloud.

A new deduplicationsystem supporting differential duplicate checkis proposed under this hybrid cloud architecture wherethe S-CSP resides in the public cloud. The user is onlyallowed to perform the duplicate check for files markedwith the corresponding privileges.Furthermore, we enhance our system in security.Specifically, we present an advanced scheme to supportstronger security by encrypting the file with differentialprivilege keys. In this way, the users without correspondingprivileges cannot perform the duplicate check. Furthermore,such unauthorized users cannot decrypt theeven collude with the S-CSP. Security analysisdemonstrates that our system is secure in terms of thedefinitions specified in the proposed security model. Finally, we implement a prototype of the proposedauthorized duplicate check and conduct testbed experimentsto evaluate the overhead of the prototype. Weshow that the overhead is minimal compared to the normalconvergent encryption and file upload operations.

ORGANIZATION:

The rest of this paper proceeds as follows. In Section 2,we briefly revisit some preliminaries of this paper. InSection 3, we propose the system model for our deduplicationsystem. In Section 4, we propose a practical deduplicationsystem with differential privileges in cloudcomputing. The security and efficiency analysis for theproposed system are respectively presented in Section 5.In Section 6, we present the implementation of our prototype,and in Section 7, we present testbed evaluationresults. Finally we draw conclusion in Section 8.

PROOF OF OWNERSHIP:

The notion of proof of ownership(PoW) [11] enables users to prove their ownership ofdata copies to the storage server. Specifically, PoW isimplemented as an interactive algorithm (denoted byPoW) run by a prover (i.e., user) and a verifier (i.e.,storage server). The verifier derives a short value $\phi(M)$from a data copy M. To prove the ownership of thedata copy M, the prover needs to send ϕ' to the verifiersuch that $\phi' = \phi(M)$. The formal security definitionfor PoW roughly follows the threat model in a contentdistribution network, where an attacker does not knowthe entire file, but has accomplices who have the file. Theaccomplices follow the "bounded retrieval model", suchthat they can help the attacker obtain the file, subject tothe constraint that they must send fewer bits than theinitial min-entropy of the file to the attacker [11].

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

IDENTIFICATION PROTOCOL:

An identification protocol _can be described with two phases: Proof and Verify. Inthe stage of Proof, a prover/ user U can demonstrate hisidentity to a verifier by performing some identificationproof related to his identity. The input of the prover/useris his private key skUthat is sensitive information suchas private key of a public key in his certificate or creditcard number etc. that he would not like to share withthe other users. The verifier performs the verificationwith input of public information pkUrelated to skU.

Fig. 1. Architecture for Authorized Deduplication

SYSTEM MODEL:

3.1 Hybrid Architecture for Secure Deduplication At a high level, our setting of interest is an enterprisenetwork, consisting of a group of affiliated clients (forexample, employees of a company) who will use the S-CSP and store data with deduplication technique. Inthis setting, deduplication can be frequently used inthese settings for data backup and disaster recoveryapplications while greatly reducing storage space. Suchsystems are widespread and are often more suitableto user file backup and synchronization applicationsthan richer storage abstractions. There are three entitiesdefined in our system, that is, users, private cloud andS-CSP in public cloud as shown in Fig. 1. The S-CSP performs deduplication by checking if the contents oftwo files are the same and stores only one of them.The access right to a file is defined based on a setof privileges. The exact definition of a privilege variesacross applications. For example, we may define a rolebasedprivilege [9], [19] according to job positions (e.g.,Director, Project Lead, and Engineer), or we may definea time-based privilege that specifies a valid time period(e.g., 2014-01-01 to 2014-01-31) within which a file canbe accessed. A user, say Alice, may be assigned twoprivileges "Director" and "access right valid on 2014-01-01", so that she can access any file whose access roleis "Director" and accessible time period covers 2014-01-01. Each privilege is represented in the form of a shortmessage called token.

Each file is associated with somefile tokens, which denote the tag with specified privileges(see the definition of a tag in Section 2). A user computesand sends duplicate-check tokens to the public cloud forauthorized duplicate check. Users have access to the private cloud server, a semitrustedthird party which will aid in performing deduplicableencryption by generating file tokens for therequesting users. We will explain further the role of theprivate cloud server below. Users are also provisionedcertificates). In this paper, we will only consider the fileleveldeduplication for simplicity. In another word, werefer a data copy to be a whole file and file-level deduplicationwhich eliminates the storage of any redundantfiles. Actually, block-level deduplication can be easilydeduced from file-level deduplication, which is similarto [12]. Specifically, to upload a file, a user first performsthe file-level duplicate check. If the file is a duplicate,then all its blocks must be duplicates as well; otherwise,the user further performs the block-level duplicate checkand identifies the unique blocks to be uploaded. Eachdata copy (i.e., a file or a block) is associated with atoken for the duplicate check.

DESIGN GOALS:

In this paper, we address the problem of privacypreservingdeduplication in cloud computing and proposea new deduplication system supporting for• Differential Authorization. Each authorized user isbable to get his/her individual token of his file toperform duplicate check based on his privileges.Under this assumption, any user cannot generatea token for duplicate check out of his privileges orwithout the aid from the private cloud server.• Authorized Duplicate Check. Authorized user is ableto use his/her individual private keys to generatequery for certain file and the privileges he/sheowned with the help of private cloud, while thepublic cloud performs duplicate check directly andtells the user if there is any duplicate. The security requirements considered in this paper liein two folds, including the security of file token andsecurity of data files. For the security of file token, twoaspects are defined as unforgeability and indistinguishabilityof file token. The details are given below.

• Unforgeability of file token/duplicate-check token. Unauthorized

users without appropriate privileges or fileshould be prevented from getting or generating thefile tokens for duplicate check of any file stored atthe S-CSP.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

The users are not allowed to collude withthe public cloud server to break the unforgeabilityof file tokens. In our system, the S-CSP is honestbut curious and will honestly perform the duplicatecheck upon receiving the duplicate request fromusers. The duplicate check token of users should beissued from the private cloud server in our scheme.

• Indistinguishability of file token/duplicate-check token. Itrequires that any user without querying the privatecloud server for some file token, he cannot get anyuseful information from the token, which includesthe file information or the privilege information.

• Data Confidentiality. Unauthorized users without appropriateprivileges or files, including the S-CSP andthe private cloud server, should be prevented fromaccess to the underlying plaintext stored at S-CSP.In another word, the goal of the adversary is toretrieve and recover the files that do not belong tothem. In our system, compared to the previous definitionof data confidentiality based on convergentencryption, a higher level confidentiality is definedand achieved.

OUR PROPOSED SYSTEM DESCRIPTION:

To solve the problems of the construction in Section 4.1,we propose another advanced deduplication system supportingauthorized duplicate check. In this new deduplicationsystem, a hybrid cloud architecture is introducedto solve the problem. The private keys for privileges willnot be issued to users directly, which will be kept andmanaged by the private cloud server instead. In this way,the users cannot share these private keys of privilegesin this proposed construction, which means that it canprevent the privilege key sharing among users in theabove straightforward construction. To get a file token,the user needs to send a request to the private cloudserver.

The intuition of this construction can be describedas follows. To perform the duplicate check for some file,the user needs to get the file token from the privatecloud server. The private cloud server will also checkthe user's identity before issuing the corresponding filetoken to the user. The authorized duplicate check forthis file can be performed by the user with the publiccloud before uploading this file. Based on the results ofduplicate check, the user either uploads this file or runsPoW.

FURTHER ENHANCEMENT:

Though the above solution supports the differentialprivilege duplicate, it is inherently subject to bruteforceattacks launched by the public cloud server, which can recover files falling into a known set. More specifically,knowing that the target file space underlyinga given ciphertextC is drawn from a message spaceS = fF1, _ _ _ , Fngof size n, the public cloud servercan recover F after at most n off-line encryptions. Thatis, for each i= 1, _ _ _ , n, it simply encrypts Fi to geta ciphertext denoted by Ci. If C = Ci, it means thatthe underlying file is Fi. Security is thus only possiblewhen such a message is unpredictable. This traditionalconvergent encryption will be insecure for predictablefile.

IMPLEMENTATION:

We implement a prototype of the proposed authorizeddeduplication system, in which we model three entitiesas separate C++ programs. A Client program is used tomodel the data users to carry out the file upload process.A Private Server program is used to model the privatecloud which manages the private keys and handles thefile token computation. A Storage Server program is usedto model the S-CSP which stores and deduplicates files.We implement cryptographic operations of hashingand encryption with the OpenSSL library [1].We also implementthe communication between the entities basedon HTTP, using GNU Libmicrohttpd [10] and libcurl [13].Thus, users can issue HTTP Post requests to the servers.Our implementation of the Client provides the followingfunction calls to support token generation anddeduplication along the file upload process.

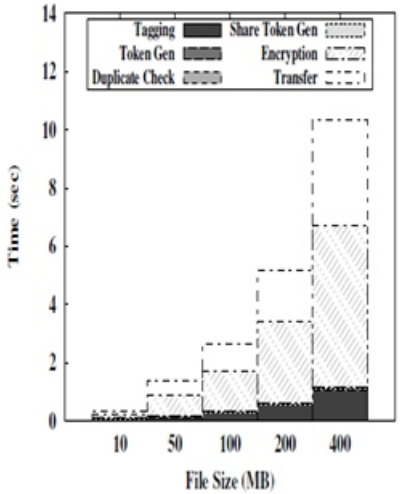

Fig. 2. Time Breakdown for Different File Size

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

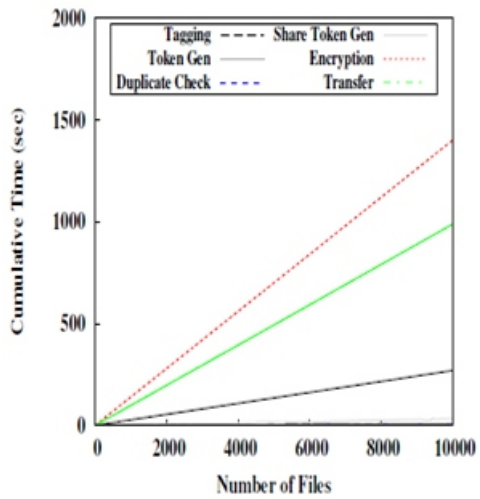

Fig. 3. Time Breakdown for Different Number of StoredFiles

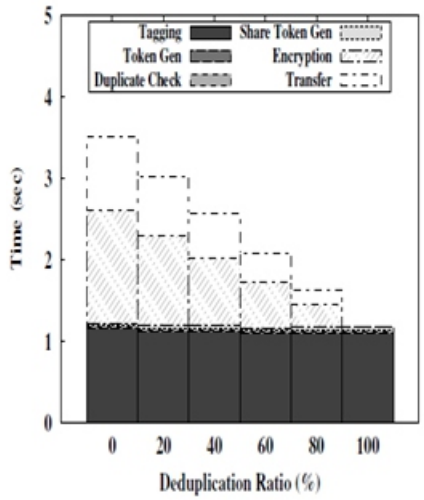

Fig. 4. Time Breakdown for Different Deduplication Ratio.

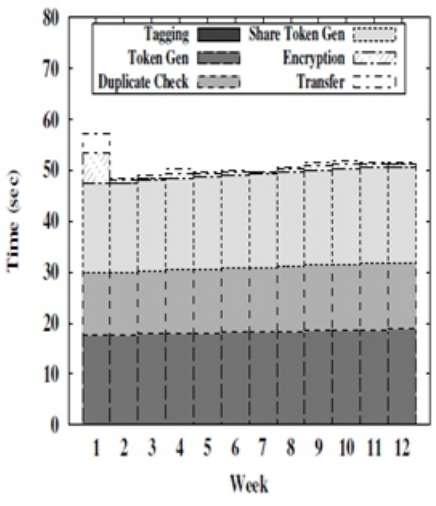

Fig. 6. Time Breakdown for the VM dataset.

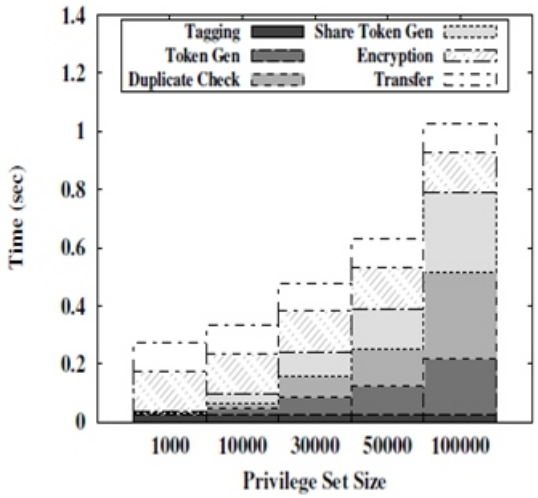

Fig. 5. Time Breakdown for Different Privilege Set Size

CONCLUSION:

In this paper, the notion of authorized data deduplicationwas proposed to protect the data security byincluding differential privileges of users in the duplicatecheck. We also presented several new deduplicationconstructions supporting authorized duplicate check inhybrid cloud architecture, in which the duplicate-checktokens of files are generated by the private cloud serverwith private keys. Security analysis demonstrates thatour schemes are secure in terms of insider and outsiderattacks specified in the proposed security model. As aproof of concept, we implemented a prototype of ourproposed authorized duplicate check scheme and conducttestbed experiments on our prototype. We showedthat our authorized duplicate check scheme incurs minimaloverhead compared to convergent encryption andnetwork transfer.

REFERENCES:

[1] OpenSSL Project. http://www.openssl.org/.

[2] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de duplication. In Proc. of USENIX LISA, 2010.

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraidedencryption for deduplicated storage. In USENIX SecuritySymposium, 2013.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

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

[5] M. Bellare, C. Namprempre, and G. Neven. Security proofs foridentity-based identification and signature schemes. J. Cryptology,22(1):1–61, 2009.

[6] M. Bellare and A. Palacio. Gq and schnorr identification schemes:Proofs of security against impersonation under active and concurrentattacks. In CRYPTO, pages 162–177, 2002.

[7] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. Twinclouds: An architecture for secure cloud computing. In Workshopon Cryptography and Security in Clouds (WCSC 2011), 2011.

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

[9] D. Ferraiolo and R. Kuhn. Role-based access controls. In 15thNIST-NCSC National Computer Security Conf., 1992.

[10] GNU Libmicrohttpd. http://www.gnu.org/software/libmicrohttpd/.